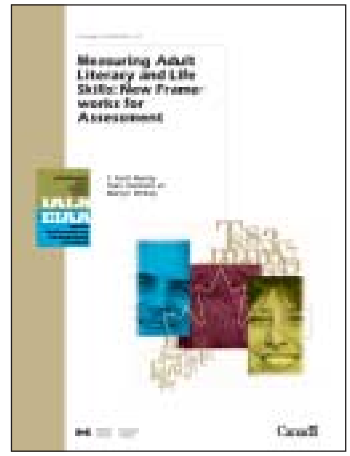
Catalogue no. 89-552-MIE, no. 13

International Adult Literacy Survey

Measuring Adult Literacy and Life Skills: New Frameworks for Assessment



T. Scott Murray, Yvan Clermont and Marilyn Binkley

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Client Services, Culture, Tourism and the Centre for Education Statistics, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7608; toll free at 1 800 307-3382; by fax at (613) 951-9040; or e-mail: educationstats@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Website	**www.statcan.ca**

Information to access the product

This product, catalogue no. 89-552-MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Measuring Adult Literacy and Life Skills: New Frameworks for Assessment

Editors
T. Scott Murray *Statistics Canada*
Yvan Clermont *Statistics Canada*
Marilyn Binkley *National Center for Education Statistics*
Washington, DC, U.S.

Published by authority of the Minister responsible for Statistics Canada

March 2005

Catalogue no. 89-552-MIE, no. 13
ISSN 1480-9516
ISBN 0-662-39834-3

Frequency: Irregular

Ottawa

Acknowledgements

This work is the results of several years of engaged participation of a countless number of individuals and organisations who participated at different development steps of these frameworks. We are therefore grateful to the numerous experts, national study managers and statistical organisations around the world who have been essential to the development and refinement of these proposed conceptual frameworks. The latest were deeply involved at each development stage from the framework review and refinement, item development and conduction of feasibility studies, data collection and analysis.

Note of Appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Table of Contents

Foreword

The objective of this report is to document key aspects of the development of the International Adult Literacy and Life Skills Survey (ALL) – it's theoretical roots, the domains selected for possible assessment, the approaches taken to assessment in each domain and the criteria that were employed to decide which domains were to be carried in the final design. As conceived, the ALL survey was meant to build on the success of the International Adult Literacy Survey (IALS) assessments by extending the range of skills assessed and by improving the quality of the assessment methods employed. This report documents several successes including:

- the development of a new framework and associated robust measures for problem solving
- the development of a powerful numeracy framework and associated robust measures
- the specification of frameworks for practical cognition, teamwork and information and communication technology literacy

The report also provides insight into those domains where development failed to yield approaches to assessment of sufficient quality, insight that reminds us that scientific advance in this domain hard won.

Nevertheless, the ALL study breaks new ground, offering a wealth of data on adult skills and their covariates. Readers are encouraged to celebrate our successes, to learn from our mistakes and to reflect on the central importance of what has been measured to the lives of our citizens.

T. Scott Murray   Yvan Clermont   Marilyn Binkley

Organization of this report

This report is organized into three parts.

Part I includes three chapters.

Chapter 1 provides readers with an overview of the policy issues that motivated the IALS, the study upon which the ALL study sought to build. The chapter identifies the explicit objectives that were set for the ALL study, the pragmatic considerations that influenced the design and documents the overall approach taken to the development and validation of the instrumentation applied in the study.

Chapter 2 introduces conceptual frameworks for understanding and assessing adult literacy and life skills. It provides readers with a summary of the DeSeCo general conceptual frame of reference for key competencies.

Chapter 3 traces development of the ALL study from the general theory that underlies the assessment through the research and development that led to final design. Readers are provided with insight into the thinking that went into the selection of skill domains, the process that was followed to develop assessment frameworks and related instrumentation in each domain, how the assessment instrumentation was validated, what criteria were established for measures to be included in the international comparative assessment of skill and where development managed to produce measures of sufficient quality.

Part II includes three chapters that provide assessment frameworks for the four skill domains that met the criteria set out for inclusion in the international comparative assessment. These chapters also document the processes that were used to develop and validate the assessment instruments and presents data related to validity and reliability.

Chapter 4 presents the framework for prose, document and quantitative literacy.

Chapter 5 presents the framework for numeracy, and,

Chapter 6 presents the framework for problem solving.

Part III includes three chapters. Chapters 7 and 8 provide assessment frameworks for two skill domains where development failed to yield approaches to measurement that were sufficiently robust to meet the criteria set out for inclusion in the international comparative assessment. These chapters also set out what was learned during the process of development and validation. Chapter 9 provides the assessment framework developed by Educational Testing Service (ETS) for measuring information and communication technology literacy. The ETS ICT literacy framework only became available after the final design of the ALL study was set so no direct assessment of this skill domain could be included. Sufficient development and validation has since taken place to confirm the viability of the framework and an approach to measurement.

Chapter 7 presents the framework for teamwork

Chapter 8 presents the framework for practical cognition

Chapter 9 presents the framework for information and communication technology literacy

Part I

Policy Interest and Theory Behind the Adult Literacy and Life Skills Project

This first part of this publication provides readers with an overview of the policy issues that motivated the IALS, the study upon which the ALL study seeks to build. It identifies the explicit objectives that were set for the ALL study, the pragmatic considerations that influenced the design and documents the overall approach taken to the development and validation of the instrumentation applied in the study. The following portions of Part 1 (chapters 2 and 3), trace development of the ALL study from the general theory that underlies the ALL assessment through the research and development that led to final design. Readers are provided with insight into the thinking that went into the selection of skill domains, the process that was followed to develop assessment frameworks and related instrumentation in each domain, how the assessment instrumentation was validated, what criteria were established for measures to be included in the international comparative assessment of skill and where development managed to produce measures of sufficient quality.

Chapter 1

The Origins and Objectives of the ALL Study

T. Scott Murray *Statistics Canada*
Yvan Clermont *Statistics Canada*

Table of Contents

1. Background

In 1994, nine countries (Canada, Germany, France, Ireland, the Netherlands, Poland, Sweden, Switzerland, and the United States) fielded the International Adult Literacy Survey (IALS), the world's first large-scale, comparative assessment of adult literacy. In December 1995, Statistics Canada and the OECD published *Literacy, Economy and Society: Results of the First International Adult Literacy Survey* (OECD and Statistics Canada, 1995), a report that presented data for seven of the countries that participated in the first round of the IALS data collection.

Encouraged by the IALS success, five countries (Australia, the Flemish community in Belgium, Great Britain, New Zealand, and Northern Ireland) decided to administer the IALS instruments in 1996. Data from this round of collection was released in November 1997 in *Literacy Skills for the Knowledge Society: Further results of the International Adult Literacy Survey* (OECD and HRDC, 1997).

Finally, nine countries (Chile, the Czech Republic, Denmark, Finland, Hungary, Italy, Norway, Slovenia, and Switzerland) participated in a third, large-scale round of data collection in 1998. Data for 22 countries was published by Statistics Canada and the OECD as *Literacy Skills for the Information Age: Final Report of the International Adult Literacy Survey* (2000).

Since that time Japan, Malaysia, Portugal, Ontario, China and Vanuatu have also successfully collected data with instruments derived from IALS.

IALS provided previously unavailable information on the distribution of adult literacy and numeracy skills and has provided tantalising insight into the causes and consequences of these skills for a range of countries.

Key findings include:

1) Important differences in literacy skills exist across and within nations, differences that are much larger than suggested by differences in national educational attainment profiles.
2) Literacy skill deficits are not only found among marginalised groups, but affect large portions of the entire adult population.
3) Literacy is strongly correlated with life chances and use of opportunities, both social and economic.
4) Literacy is not synonymous with educational attainment.
5) Literacy skills, like muscles, are maintained and strengthened through regular use.
6) Adults with low literacy skills do not usually acknowledge or recognize that their skills may pose a problem.

Not surprisingly, the IALS reports attracted a great deal of interest from national policy makers and the popular press. It is clear that the study has answered many questions of pressing interest and concern. Yet, as with any well-conceived study, IALS raised many new questions. Key among such questions are those that speculate about the relationship of literacy skill to other skills thought to be important to workforce productivity and labour market success. Many studies, including the Secretary's Commission About Necessary Skills (SCANS) in the United States, have posited the existence of a range of skill domains thought to be of economic importance. Yet little, if any, empirical evidence exists to test these notional skill frameworks.

The success of the IALS approach led several national governments to wonder if the methods could be adapted to measure a broader array of skills on an international level.

A first meeting to consider the possibility of mounting such a study was hosted by the Swedish Educational Authority Skolverket. Documentation prepared by Statistics Canada for consideration[1] reviewed the prevailing notions of basic skills and offered a hybrid typology with eight distinct domains which might be included in an IALS-type study (Jones, 1996). Statistics Canada suggested that the proposed study would, in each of the eight skill domains, administer a computer-based test to a nested sample of workers within firms so that explicit statistical linkages would be available to isolate the impact of observed skill on economic productivity and indicators of firm success such as employment growth and profitability.

Interest in this idea was sufficient amongst national governments to organise a second meeting, hosted by the University of Amsterdam, to discuss the merits of available conceptual frameworks and to review the validity, reliability and operational feasibility of related assessment technology. The meeting concluded that coherent conceptual frameworks and satisfactory measurement technology did indeed exist for several, but not all, of the proposed skill domains. It became clear, however, that the costs and operational implications of fielding a computer-based test to a nested sample of workers within firms were well beyond the financial and technical capability of many of the prospective participants.

This realisation led to a meeting of the International Study Team to consider options. Hosted by the US National Center for Education Statistics in Washington DC, the meeting concluded on pragmatic grounds that:

1) The proposed assessment should use paper and pencil rather than computer-based tests.
2) The development of frameworks and associated assessment methods should be limited to six of the initial eight skill domains:

   prose literacy,

   document literacy,

   numeracy,

   teamwork,

   problem-solving,

   practical cognition, and,

   working with information technology.

   While of interest, other skill domains lacked either satisfactory theory or measures deemed viable within the context of a paper and pencil household-based assessment.
3) The test should be administered to representative samples of adults drawn from households rather than from workers within firms.

1. See Background for Canadian Basic Job Skills Test, S. Jones, Human Resources Development Canada and Statistics Canada, 1996.

These basic planning assumptions were subsequently ratified at both the first and second meetings of the ALL Project Advisory Group (PAG) which is responsible for providing management oversight and advice to the Project within the OECD programme of work.

A meeting of experts, convened in May 1997, reviewed the proposed design of the study and the frameworks for each of the skill domains. Held at the OECD in Paris, the meeting concluded that additional work needed to be done on several of the skill frameworks if the study was to have a reasonable chance of generating valid, reliable and comparable skill profiles at the international level. As a result, the teams responsible for numeracy, problem-solving, teamwork and practical cognition were funded to refine their frameworks and to collect sufficient empirical data to demonstrate the measurement properties of the proposed measures trans-nationally.

Two subsequent meetings of the international study team were hosted by the US National Center for Education Statistics in Washington, DC to review progress and plan for further development. Held in April 1998 and September 1998, these meetings concluded that while the proposed frameworks for problem-solving and attitudes towards teamwork were judged to be adequate, the approach to measurement had failed to yield data of sufficient quality. In addition, the proposed instrumentation for measuring computer literacy was judged inadequate. As a result, new development teams were recruited and funded by Statistics Canada, NCES and the Governments of Sweden and Luxembourg.

Two additional meetings were held.

- First, a meeting of all development team members was held in Washington, DC January 20–23, 1999 to help integrate the different assessments and to provide expert feedback.
- A second meeting of development team leaders was held in Princeton, NJ on August 23–24, 1999 to review the frameworks and the results of small-scale, multi-country piloting with a view to selecting domains for inclusion in the ALL pilot.

International teams were subsequently recruited to draft final assessment frameworks and test specifications and to develop items in seven cognitive domains:

- prose literacy
- document literacy
- numeracy
- problem solving
- practical cognition
- teamwork and,
- information and communication technology

A separate international team developed the background questionnaires. In total representatives of some 35 countries, representing 20 languages, participated in the item writing activity.

Small scale piloting, in pairs of countries, failed to yield measures of sufficient reliability for:

- practical cognition
- teamwork, and,
- information and communication on technology

Measures of attitudes to teamwork and familiarity and use of ICT's were incorporated into the ALL Pilot background questionnaires but only the questions pertaining to the familiarity and use of ICT were retained for the main ALL study.

Large scale piloting by countries participating in the first round of ALL collection provides evidence that the ALL instruments yield valid, reliable, comparable and policy relevant data.

Six countries – Bermuda, Canada, Italy, Norway, Switzerland, and the United States – participated in the first round of ALL data collection, fielding the ALL pilot study in 2002 and the main data collection in the first and second quarters of 2003. The Mexican state of Nuevo Leon fielded, in 2003, a hybrid assessment that employed the IALS assessment and the ALL background questionnaire.

2. Objectives

The ALL study is, by design, meant to inform public policy in a number of related areas including education, labour market policy, human resource development and social development. The policy aims and objectives of the ALL study are extensively documented by Giddings and Barr-Telford (Statistics Canada 2000).

Put succinctly the Adult Literacy and Life Skills (ALL) Survey was designed to serve several sets of related objectives.

First, the study hoped to:

- Profile the distribution of prose literacy and document literacy in the adult population.

One of the key policy concerns raised through analysis of IALS data was the putative existence of significant levels of skill loss in adulthood. In response the ALL study was designed to shed light on the social and economic factors that determine or underlie change in the observed skill profile over time. In order to shed empirical light on this concern, the prose literacy and document literacy domains to be measured in ALL will be linked to the original IALS scales through the administration of common items. For those countries that fielded IALS, such item-level linkage will provide a powerful means of exploring trends in the evolution of skill profiles. This information is crucial for judging the relative priority of policies related to basic adult skills and for identifying where policy might best focus itself – on skill demand, on skill supply, on the efficiency of markets for skill or on some combination of the three.

Second, the study sought to:

- Profile the distribution of numeracy, and problem-solving, in the adult population, and determine the inter-relationship of these skills to prose literacy and document literacy.

The IALS data collection has yielded an empirical data set of some 100,000 observations, providing researchers with a substantial amount of information with which to explore the social and economic causes and consequences of the observed skill data.

A basic objective of the ALL will be to build on this understanding by adopting a design which will allow an empirical appreciation of how performance on each of the newly tested skills relates to the skills tested in the original IALS study. The rapid rate of technological innovation and the globalization of markets have led to high rates of structural adjustment in many OECD economies. This, in turn, has quickened the pace at which disadvantaged individuals become marginalised. In many cases, the very structure of educational systems and the labour market work against rapid adjustment and the interests of the marginal workers because the systems for signalling skill seek to divide the workforce into discrete, non-transferable categories. The ALL seeks to empirically establish the existence of generic skill clusters that transcend industry, occupation, educational qualifications and age-based experience.

Thirdly, the ALL study attempted to:

- Determine the relationship of each of the tested skills to individual economic and social success.

Many studies have documented the relationship of educational attainment to social and economic success but, until IALS, few studies had sufficient data to allow the empirical exploration of how this relationship depended, in turn, on more fundamental processes such as actual demonstrated skill. IALS has revealed that literacy and education are not synonymous, and that social and economic success depends, in part, on tested skill. The study has also revealed interesting variation in these relationships both within occupations and between countries, facts that fit with common wisdom about the impact of economic and social organisation on markets for skill. The ALL study allows for an extension of this basic analysis to additional skill domains, and where IALS data are available, an understanding of how these relationships are evolving as the knowledge economy develops.

Fourth, the ALL study intended to:

- Identify sub-populations whose performance places them at risk.

Much of the rhetoric employed in the North American debate about skills has been focussed on their impact on the so-called "high performance" workplace. Despite this fact, much of the attention of governments continues to be directed towards those groups of people whose skill levels place them at risk of being socially and economically marginalised. IALS has revealed that, in some economies at least, individuals with poor skills experience significant wage and employment penalties. The design of policies and programmes to attenuate the worst of these impacts and to provide remedial education depend entirely on understanding the number, geographic distribution, and characteristics of the people so affected. The ALL study attempts to profile those whose performance places them at risk.

The ALL also hoped to meet a number of longer-term objectives, including:

- To shed light on the causes and consequences of the observed skill distributions

Longitudinal data is needed in order to truly understand the causes and consequences of any human phenomena, as it is only longitudinal data that allows one to disentangle cohort, life-cycle, and period effects. Thus the ALL, conceived to provide a cross-sectional "snapshot" of the distribution of skill in several domains, cannot be expected to advance our understanding in this regard. When analysed in conjunction with key co-variates provided by the background questionnaire, however, the ALL can be expected to yield tantalising evidence about the relative impact that the various factors might have on the observed distributions of skill. This will be particularly so in those countries where IALS data has already been collected. In this case, changes in the distribution of prose and document literacy skill between the two observations can be related, in a synthetic-longitudinal analysis, to changes in the underlying co-variates. Given the nascent state of skill measurement at the population level, the high cost of longitudinal surveys, the length of time it takes a longitudinal study to yield data, and the horrendous cost of measuring the wrong things on such a study, the ALL can provide critical data to inform the next generation of true longitudinal studies, including the design of Canada's Youth in Transition Longitudinal Survey program.

- To contribute to the literature on the basis of human cognition

The IALS study was based on a powerful theory of adult reading, a theory that served as a basis for both the design of the assessment and the interpretation of the resultant data. Thus, IALS afforded researchers a unique opportunity to empirically validate the theory with data of unparalleled coverage and large, representative samples of adults. The ALL offers a much greater potential to shed empirical light on the validity of several of the various competing frameworks which deal with the organisation and structure of human intelligence and cognition.[2] A long term objective is, therefore, to assure that the ALL data is made available to researchers for this purpose, particularly those researchers involved in the ongoing work of the OECD's DeSeCo project.

- To foster continued international co-operation on the design, implementation and analysis of data on the distribution and co-variates of skill

Direct measurement of the sort employed in the IALS study requires considerable operational and technical skill and significant financial resources to unilaterally design, validate, collect and analyse. Such resources are beyond the means of many of even the most advanced economies to support. IALS has also demonstrated the potential of a comparative perspective to shed light on deep relationships underlying the observed phenomena, relationships that remain undetected in idiosyncratic national studies.

2. See for example, Toward a triarchic theory of human intelligence, Robert J. Sternberg, 1984 Cambridge University Press, 1984 and Frames of Mind: The Theory of Multiple Intelligences, Howard Gardner, Harper Collins, 1993.

3. Development and validation of assessment frameworks and instruments

Development and validation of the ALL assessment frameworks and assessment instruments has been an expensive and lengthy process. The overall goal of this process was to create instruments that are:

valid,

reliable,

comparable,

interpretable,

policy relevant,

amenable to policy intervention, and,

linguistically, culturally and geographically appropriate.

Development of assessment instruments for the ALL study was attempted in 7 domains:

1. Prose literacy
2. Document literacy
3. Numeracy
4. Problem solving
5. Practical cognition
6. Teamwork
7. Information and communication technology

ALL development proceeded in six distinct phases:

- development of an overarching framework to define skill domains where measurement should be attempted;
- development of assessment frameworks, test specifications and exemplar assessment items.

In each case an international panel of experts conducted a review of existing theory and approaches to measurement. This review was used to construct a framework for assessment that rendered explicit the factors that underlie the relative difficulty of tasks in each domain.

- small scale piloting in pairs of countries to confirm key theoretical and measurement assumptions;
- large scale development of assessment items in each domain by extensive networks of international experts.

In each case a broader group of experts was enlisted to use the assessment frameworks to develop assessment items that covered the expected range of proficiency and social contexts.

The consortium went to great lengths to ensure that development was broadly representative linguistically, culturally and geographically.

In a parallel activity assessment frameworks were refined to reflect what was learned through small scale piloting and item pools were bundled into a pilot survey for testing in all participating countries.

- full scale piloting of ALL instruments in seven countries and five languages;
- selection of items for inclusion in the final assessment, refinement of the frameworks and background questionnaire using data from the full pilot.

4. Criteria for inclusion of tests assessments in ALL

Having identified the domains in which the ALL study would attempt to develop assessments strict criteria were established for inclusion in the final international comparative assessment.

In keeping with the initial selection criteria, skill domains carried in the international assessment had to be related to key health, educational, social or economic outcomes. At this stage an additional domain had to explain at least an additional 10% of variance of at least one of these outcomes.

Second, the theory in each domain had to identify a set of variables thought to underlie the relative difficulty of tasks in the domain, a set that, a priori, had to theoretically explain most of item difficulty over the intended range of assessment described by the framework.

Third, empirical results had to demonstrate a high degree of agreement between item difficulties predicted from theory and those estimated empirically from pilot data. For the ALL study the agreement rate had to exceed 80% at the population level.

Fourth, empirical pilot results had to demonstrate that items were working in a psychometrically stable and equivalent for population sub-groups within countries and between countries.

Inter-country comparisons of percent correct, omit rates, not reached rates, biserial correlations and item response theory (IRT) parameters were examined to determine that they conformed to expected patterns. The mean deviations, and root mean square deviations of item characteristics curves were computed to ensure that items were functioning to an empirically defined tolerance both within and between countries.

Fifth, open ended items had to be scoreable to a very high reliability – within 97% or better inter-rate agreement within countries and 90% or better inter-rate agreement rates between countries – to ensure comparability.

Sixth, estimates of internal consistency of the test had to display an average r-biserial correlation in the range of 0.60, a level that assures that items are reliably measuring a same and single underlying dimension.

Seventh, assessment items needed to take little enough time to allow each respondent to take multiple items, a prerequisite to good statistical coverage of the construct and its covariance with background characteristics.

Eighth, assessment items had to cover the range of proficiency demonstrated by 95% of the target populations, thus assuring that there is no ceiling, nor floor effect.

Ninth, items had to discriminate proficiency over the range of difficulty/proficiency displayed by the bulk of the population. In addition items needed to display good psychometric properties, particularly with respect to the stability of fit across languages and cultures and reasonably steep slopes.

Tenth, assessment items had to be culturally diverse, representing a broad range of cultures, languages and geographic regions.

Similarly, items had to be sampled from multiple life contexts i.e. home and family, health and safety, community and citizenship, consumer economics, work and leisure and recreation so that individuals of all walks of life were able to find familiar tasks, and no single group would be placed at a disadvantage. Items also had to systematically differ in the type of thinking involved and the type of materials used, elements specified in each framework.

Finally, the assessment methods used had to minimize the amount to which items in a particular domain were dependant on skill in other domains. That is especially true for designing items of low level of difficulty in traits where pure orthogonality between domains is difficult to assume. For example, steps were taken to reduce the amount of reading associated with low level numeracy items so that individuals with low reading skills would have an opportunity to display their numeracy skills.

Development and validation of the ALL frameworks and assessment protocols

Armed with these criteria, a four phase research and development program was launched.

First, an international team of experts was recruited to review the relevant literature with a view to:

- defining the domain
- specifying the variables that underlie relative difficulty
- developing example items to illustrate the recommended approach to assessment

The assessment frameworks and test blueprints produced by these teams were exposed to critical review and revised accordingly.

Second, a variant of each assessment was piloted in at least two countries in order to establish that:

- the items functioned as predicted, and
- respondents reacted in a positive way to the assessment items

At this point the ALL developers faced a decision.

Frameworks and associated approaches to measurement that were judged to be sufficiently robust were referred on to a phase of international item development designed to:

- expose a large number of researchers to the theory and approach to measurement proposed for ALL
- ensure that items selected for inclusion in ALL provided good linguistic, geographic and cultural coverage.

The frameworks for prose literacy, document literacy, and numeracy met the criteria established for inclusion and were referred to international item development. Chapter 4, 5 and 6 document the frameworks, the item development process and how the assessment protocols were validated at the international level. The initial Problem Solving framework and approach to assessment proposed by Harry O'Neil and this team at CRESST, UCLA proved to be unreliable in initial multi-country piloting and a second group of experts, headed by Jean-Paul Reeff, was recruited to refine the framework and to develop on an alternative approach to assessment. This redevelopment, effort, documented in Chapter 6 of this report, proved successful.

Chapter 7 in Part III also documents the development effort for teamwork headed by O'Neil and his colleagues at CRESST. It produced a useful framework but failed to provide an approach to assessment of sufficient quality to merit inclusion in an international comparative study. A second team, headed by David Baker at the America Institutes for Research, developed a slightly modified framework and a different approach to assessment. This effort also failed to meet the standards set for inclusion in ALL. Despite failing to yield the desired result – a valid, reliable comparable and interpretable teamwork scale – the development effort did provide some useful results.

Analysis of small scale pilot results suggests that teamwork is a complex multi-level phenomenon with effects at the individual, the occupational, the workgroup, the firm, the industry and the national level as well as a temporal aspect that depends on where the firm is in its product lifecycle and corporate strategy. To be informative, sampling strategies must reflect all of these levels explicitly, something that is beyond the scope of the current ALL household-based design.

The framework and associated measures for practical intelligence have been included as Part III of this report (Chapter 8). Although these were deemed to be adequate the statistical techniques employed to summarize proficiency were judged to be difficult to interpret within the context of an international comparative study. This domain was dropped from the international study pending additional work.

Development of a framework for measuring information and communication literacy, headed by Jean-Paul Reeff, then of the Luxembourg Ministry of Education, failed to provide a clear approach to measurement. As a result Graham Lowe of the University of Alberta was recruited to develop a behavioural module, focusing on familiarity and use of ICT's, to replace a direct measure of ICT proficiency. This module was eventually included in the ALL assessment. In a parallel activity Statistics Canada chose to invest in an ETS project designed to develop a framework and viable approach to measurement of ICT skills. Although this development failed to yield results in time for inclusion in the ALL study it did establish the basis for doing so in future assessment cycles. Interested readers are referred to chapter 9 for additional detail on this work.

5. Development of the ALL Background questionnaire

The ALL Background Questionnaire (BQ) was designed through a collaborative effort by members of the International Study Team, National Study Managers (NSMs), and individual experts. The final version of the questionnaire was the result of extensive consultation, testing, qualitative and quantitative evaluation of the pilot survey instrument. The overall development of measures of adult education and training benefited greatly from work done by the PEL sub-group of the OECD's INES network (led by Anna Borkowsky of Switzerland). The numeracy assessment team, lead by Iddo Gal, also consulted on numeracy-oriented measures included in the BQ. Westat in the United States led the work on determining appropriate measures of health and well-being. Two subgroups of NSMs, each representing six countries, worked on the initial design of adult education indicators and social capital indicators. Expert consultations were on-going with Stan Jones. The questionnaire benefited from comments from Albert Tuijnman, Emmanuel Boudard, Tom Healy, Paul Reed, and Gordon Darkenwald.

The BQ process began at the first National Study Managers Meeting in September of 1999. In preparation for this meeting, Westat submitted a series of recommended changes to the IALS background questionnaire and several experts' suggestions for content development/improvement were included in the meeting documentation package. Based on these recommendations, a first draft BQ was designed for discussion at a major BQ conference held in March 2000 in Ottawa, Canada. NSMs and experts gathered to discuss the proposed BQ content. At this conference, the structure of question types - internationally required, internationally optional, and nationally required - was agreed upon, several questions were designated to those categories, international coding structures were determined and each module was reviewed in detail. Two working groups (representing 6 different countries in each group) were formed to improve the content in two domains – 'Participation in Education and Learning' and 'Social Capital'. Following the conference, a table outlining all proposed changes and recommendations for the BQ was circulated for verification and approval. A second draft BQ was created based on the comments and recommendations made at the international BQ conference. This draft was then tested in Canada and Italy. Based on the testing results, the International Study Team recommended a series of adjustments to the BQ that were presented for comment and discussion at the NSM meeting held in June 2000 in Frascati, Italy. A third draft BQ was created after the NSM meeting incorporating many of the International Study Team recommendations and comments from NSMs. This draft was circulated for comments in July 2000 and a final BQ for pilot testing was produced.

6. Summary

As noted above the ALL study was designed to respond to a range of important public policy issues related to adult education, skills and learning.

To fulfill these goals Statistics Canada and US National Center for Education Statistics funded an ambitious program of research and development designed to:

- identify a set of skill domains grounded in theory and thought to be related to success in life and to a well functioning economy and society
- develop approaches to measurement that afford valid, reliable, comparable and interpretable profiles of skill for heterogeneous populations within and between countries, within the natural constraints of a household survey of adults
- associate these skill profiles with a range of background variables designed to reflect the social distribution of skill, the factors that influence the level and distribution of skill, the health, social, educational and economic outcomes that are associated with different levels of skill at the micro, meso and macro level, and individual's own assessment of his/her skill and its relationship to their economic and social success.

  In keeping with the importance of the issues at play, the ALL survey set demanding scientific standards for inclusion in the final international comparative assessment.

Development failed to realize measures of sufficient quality in three domains:

- practical intelligence
- teamwork, and,
- information and communication technology

As a result the ALL survey provides valid, reliable comparable and interpretable profiles of prose literacy, document literacy, numeracy and problem solving.

The profiles for prose literacy and document literacy will be linked at the item level to permit an analysis of the evolution of the skill profile since 1994 and to identify the factors that seem to have the greatest impact on change (e.g. skill gain, and/or skill in adulthood or the quality and quantity of initial education).

The design will also permit the estimation of the inter-skill covariance matrix information critical to devising efficient and effective remedial education programs.

An attempt will also be made to link the quantitative literacy domain of the IALS study to the broader ALL numeracy assessment.

The ALL study will also provide:

- an index of respondent familiarity with and use of information and communication technologies. This data will be used to explore the degree to which such skills depend, in the first instance, on high levels of literacy and/or numeracy skill and the extent to which they amplify inequality in social economic, educational and health outcomes.
- A profile of respondent's participation in formal and informal adult education and training, one of the key suspected mechanisms of skill retention/acquisition.

- The inter-skill covariance matrix that reveals the degree to which skill in one domain depends on skill in other domains. Such dependencies have important implications for the structure and content of remedial education.
- The relationship of skill to a broad range of demographic variables to identify groups whose skill either places them at risk or provides significant advantages in various life contexts.
- The relationship of skill to a broad range of variables thought to influence the acquisition and maintenance of skills, including, as noted above, their participation in adult education and training.
- The relationship of skill in each domain to a range of objective economic, social, educational and health outcomes.
- Indices of respondent's use of skill at work and other life contexts.
- Measures of respondents' perceptions of their own skill levels and the degree to which they believe that their skill acts as a barrier to their life chances.

The ALL study was designed to yield data of unparalleled quality on a range of public policy issues broadly related to skills and learning. The research and development effort associated with the study has provided the scientific basis to support this goal.

Experience with assessment at the international level suggests that this is a necessary, but not a sufficient, step in ensuring valid, reliable comparable and interpretable results. Quality depends, ultimately as much on how the study is implemented as on the design.

References

Jones, S. (1996). *Background for Canadian Basic Job Skills Test*, Human Resources Development Canada and Statistics Canada.

Giddings, D. and Barr-Telford, L. (2000). *Skill development and public policy*, Statistics Canada and National Center for Education Statistics (unpublished paper).

OECD and Statistics Canada, (1995). *Literacy, Economy and Society: Results of the First International Adult Literacy Survey.*

OECD and Human Resources Development Canada, (1997). *Literacy Skills for the Knowledge Society: Further results of the International Adult Literacy Survey.*

Statistics Canada and the OECD, (2002). *Literacy Skills for the Information Age: Final Report of the International Adult Literacy Survey*.

Sternberg, R.J. (1984). *Toward a triarchic theory of human intelligence*, Cambridge University Press, 1984 and Frames of Mind: The Theory of Multiple Intelligences, Howard Gardner, Harper Collins, 1993.

Chapter 2

Conceptual Frameworks for Understanding and Assessing Adult Literacy and Life Skills

Dominique Simone Rychen *Swiss National Statistics Office*
T. Scott Murray *Statistics Canada*

Table of Contents

1. Introduction

Few people would argue against the idea that there are skills and abilities necessary for success in life. Asking people to name them, however, would generate a wide variety of responses. This should not be surprising, since skills and abilities important to one person may not be as important to another. Differences may arise from occupation (e.g., corporate executive vs. assembly line worker), lifestyle (e.g., head of a large household vs. single with no dependents), society and culture (e.g., industrialized vs. agrarian) or from differences in the dominant technologies of production and associated ways in which work is organized (eg.Tayloristic production versus high performance work groups). Despite these differences, there has been a great deal of interest in trying to look across individual and cultural contexts to identify and measure a common, definable core of necessary skills and abilities. This is where the Adult Literacy and Life Skills survey (ALL) begins—as an attempt to identify and measure a range of skills that are linked to social and economic success with the goal of developing profiles that capture variations across groups and the environments they in which they operate.

Understanding these empirical linkages is important for both public policy and individual choice (OECD and Statistics Canada 1995; OECD and HRDC 1997; OECD and Statistics Canada, 2000). First, skill is thought to be an important force driving both aggregate economic performance and inequality in educational, social and economic opportunity at the individual level. Skill is also thought to play a central role in the generation of and access to social capital (Bourdieu, 1977 and Coleman, 1988), and to support the development of, and access to, democratic institutions (Freire, 1970).

The designers of the ALL study did not begin with a blank slate. Recently, there has been a proliferation of efforts in the fields of education and labour to develop lists of skills, knowledge, and competencies necessary for success in the workplace and society. Thus, the effort could be as simple as reviewing these studies to identify the one that is most appropriate, or a set of skills common to most of them. As appealingly straightforward as this sounds, this body of research is not the only one relevant to this purpose. Indeed, over the past century, researchers from a variety of fields have sought to identify models and systems to describe concepts very similar to, if not the same as, life skills. Most prominent among these is the work that has been done to define human intelligence. Because current notions of intelligence extend well beyond academic knowledge, one might at first expect them to resemble the sets of employability skills identified by the education and labour researchers. Examining the two together, however, reveals striking contrasts in approaches and language. For example, predominant workplace skill models frequently originate from inventories of tasks encountered in jobs and everyday situations. On the other hand, models of intelligence seek to identify products and processes of abilities, referring to skills and tasks only as a means of exemplifying and measuring these abilities, or as subcategories of them.

Having examined work in both of these fields, the ALL designers identified the need for two related strands of conceptual research and development.

One strand, known as the DeSeCo project, responded to the need to embed the ALL study skill measures within a broader conceptual framework. Such broad conceptual frameworks serve three important purposes:

- they identify the skill domains that should be considered for assessment,
- they identify the non-cognitive dimensions that are associated with competent performance, and
- they help to guard against the over-interpretation of what can be measured with current large scale assessment technology at the expense of what can not yet be measured.

The second strand of work attempted, through a process of formal analysis, to move towards measurement by combining notions of skill embodied in the literature on human intelligence with those derived from the workplace skill literature.

The next section of this report provides a summary of the DeSeCo project and its conclusions. This is followed by an overview of the work that was undertaken to define a more focussed overarching framework to guide the development of valid, reliable, comparable and interpretable measures.

Several general conclusions can be drawn from these two strands of work:

- The perspectives developed in DeSeCo and ALL complement rather than contradict each other
- The skills measured in the IALS study – prose literacy, document literacy and quantitative literacy – were identified as critical elements to be measured in both perspectives
- The ALL study should seek to extend measurement to additional skill domains where sufficient theory and approaches to measurement warranted.
- The instruments employed to assess skill should be sensitive to context and incorporate aspects of the psycho-social prerequisites of competence.

2. DeSeCo: A general conceptual frame of reference for key competencies

DeSeCo - the acronym of *Definition and Selection of Competencies: Theoretical and Conceptual Foundations* – was an interdisciplinary, theory-oriented study that originated in the context of international efforts to assess education and learning outcomes and to evaluate educational systems.

The DeSeCo project was led and managed by the Swiss National Statistics Office (BFS), financed by the Swiss, Statistics Canada and the US National Center for Education Statistics and implemented under the aegis of the OECD. The project was launched in late 1997 in recognition of the need to conceptually bridge past and current international competence assessments and to develop an overall long-term strategy for international assessment of competencies based on a common theoretical and conceptual foundation.

Which competencies are relevant for individuals to lead a successful and responsible life and for society to face the challenges of the present and future? Is it possible to determine a limited set of key competencies important in different spheres of life, such as the economic sector, the political sphere, and family life? And if so, what is the nature of these competencies? These are the type of questions that guided DeSeCo's research process.

The multiple inputs to DeSeCo and the exchange among a wide range of experts and stakeholders resulted in a comprehensive frame of reference for key competencies.

This section lays out DeSeCo's mission and research process, followed by a succinct presentation of the core elements of the frame of reference elaborated in the final report (Rychen and Salganik, Eds. 2003). It concludes with a brief discussion of its relevance for ALL and of the challenges for future international competence assessments.

3. DeSeCo and ALL are complementary research efforts in response to policy needs

The DeSeCo Project was designed as a complementary theory-oriented study to international large-scale assessments such as TIMSS, IALS, ALL and PISA and is part of a common effort to respond to an increased policy interest in information about the quality and adequacy of education and training, the role of competence in creating social inequity in individual economic, social and educational outcomes and the effects of human capital investment on overall levels of social and economic development.

As noted above ALL defined life skills[1] based on the scientific literature on cognition and on occupation skill standards and job analyses and measured relevant cognitive skills that correspond to the demands of the labour market and the broader society thought to be critical to economic and social success, while DeSeCo - although taking these important aspects into account (e.g. Levy and Murnane, 2001; Weinert, 2001; Kegan, 2001; Canto-Sperber and Dupuy, 2001) - sought to approach the question of key competencies from an interdisciplinary and policy perspective.

DeSeCo's main goal was to develop theoretical and conceptual foundations for defining a comprehensive set of key competencies that are relevant for personal, social, and economic well-being.

Different from ALL, considerations of how the identified competencies can be assessed or measured were not at the forefront of DeSeCo.

Yet, from the outset, close linkages with ALL and other international large-scale assessments were maintained in recognition that success in this complex field depends not only on theoretical *and* empirical work, but also on a constant *dialogue and exchange* among the various specialists and stakeholders to assure that an iterative process takes place (e.g. Murray, 2003; Schleicher, 2003).

Together, DeSeCo and current comparative assessments such as ALL and PISA, provide theoretical and empirical evidence that it is worthwhile to invest in key competencies for all as a means to enhance sustainable socio-economic and democratic development of societies.

1. The term "key competencies" as defined by DeSeCo would actually better reflect its meaning.

4. An interdisciplinary approach to key competencies

Starting from the assumption that identifying key competencies calls for scientific discussion and analysis, but implies practical considerations and political negotiations as well, DeSeCo opted for an interdisciplinary and policy-oriented approach. A wide range of stakeholders were engaged in the process of defining and selecting key competencies at the international level. DeSeCo involved and linked sociologists, economists, anthropologists, philosophers, psychologists, a historian, education researchers, statisticians, assessment specialists, policy-makers and policy analysts, unionists, employers and stakeholders representing various sectors and national and international institutions.

DeSeCo's work program consisted of several major activities:

> The project started with an analysis of international studies conducted during the 1990s in OECD countries related to indicators of education outcomes (Salganik, Rychen, Moser and Konstant, 1999), followed by a study reviewing scholarly work on the concept of competence (Weinert, 2001), and expert opinions by scholars from five different academic disciplines, who each was asked to construct a set of theoretically grounded key competencies, and comments from policy and practice (Rychen and Salganik, Eds. 2001).
>
> A first international symposium in 1999 brought together academics and stakeholders from various fields. The conclusions from these and subsequent discussions represented a first step towards interdisciplinary insight. Then, a country consultation was organized within the OECD to review national experiences in defining, developing, and assessing key competencies (Trier, 2003). The second international symposium provided an opportunity for working towards a consensus on key competencies among a wide range of countries, stakeholders, and interest groups (Rychen, Salganik and McLaughlin, Eds. 2003).
>
> DeSeCo's main conclusions and recommendations were submitted in form of a strategy paper (OECD, 2002) to the relevant OECD Committees, and the elaborated findings have been published in DeSeCo's final report entitled "Key competencies for a Successful Life and a Well-Functioning Society" (Rychen and Salganik, Eds. 2003).

5. A comprehensive frame of reference for assessing and developing key competencies

Defining and selecting key competencies is not simply a theoretical and methodological question; it is also influenced by what are considered important and desired outcomes at the individual and societal level. Universally accepted objectives such as the respect for human rights, democratic values, and sustainable development can provide a common vision of society as a normative anchoring point for the discourse on key competencies.

The research conducted under the scope of DeSeCo resulted in a widely accepted definition of the concept of key competence.

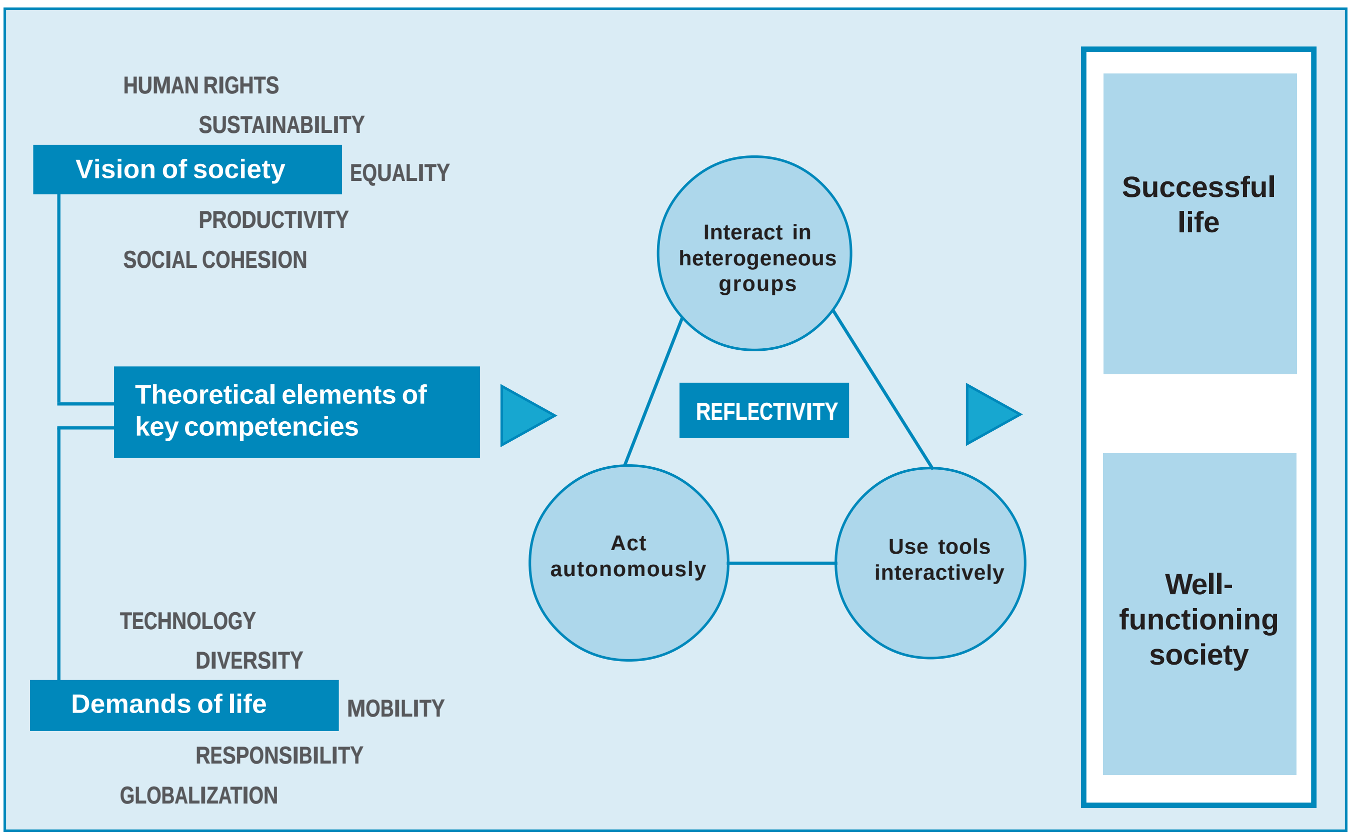
Key competencies are defined as individually based competencies that:

- are instrumental for meeting important, complex demands *in multiple areas of life*;
- *contribute to highly valued outcomes* at the individual and societal levels in terms of a successful life and a well-functioning society; and
- are important to *all individuals* for coping successfully with complex challenges in multiple areas.

In light of these definitional criteria and against the backdrop of broad societal goals and the demands of an increasingly interdependent and rapidly changing world, DeSeCo conceptualized a frame of reference for key competencies that builds on various concepts and theoretical models.

Figure 1

DeSeCo's overarching frame of reference



5.1 A demand-oriented approach to competence: the ability to meet complex demands

The definition of key competence is based on a competence model that is holistic and dynamic in the sense that it integrates and relates *demands*, *psychosocial prerequisites*, and *context* as essential elements of competent performance.

Put differently, a competence is defined as the ability to successfully meet complex demands in a particular context through the mobilization of knowledge, cognitive skills but also practical skills, as well as social and behaviour components such as attitudes, emotions, and values and motivations. This holistic apprehension is supported by recent findings from neuroscience, namely, "that reasoning and emotion are vitally connected" (Gonczi, 2003).

To illustrate, the ALL framework defines the competence "numeracy" as involving not only the enabling knowledge and specific cognitive skills required to manage the mathematical demands of diverse situations effectively, but also encompasses the activation of a range of behaviours and processes (Gal, Tout, van Groenstijn, Schmidt, and Manley, 1999).

A competence is therefore not reducible to its cognitive dimension, and thus the terms competence and skill are not synonymous. While the term *competence* designates a complex action system encompassing cognitive and non-cognitive components, the term *skill* is mostly used in relation to cognitive or practical abilities.

5.2 Reflectivity - the required level of competence

Today, in most OECD countries, value is placed on entrepreneurship and personal responsibility. Adults are expected to be flexible, adaptive, innovative, creative, self-directed, self-motivated, and to take responsibility for their decisions and actions as learners, workers, citizens, family members, or consumers.

There is a broad consensus among scholars and experts that recalling accumulated knowledge, abstract thinking, and being well-socialized are necessary but insufficient psychosocial conditions for coping with many of the demands of modern life.

Dealing flexibly with novelty, change, and diversity and coping in a responsible way with many of today's challenges require the development of a higher level of mental complexity that implies critical thinking, creative abilities and a reflective practice (including metacognition).

The notion of reflectivity (though various terms are used) is like a leading thread throughout the various contributions from scholars and experts (see in particular, Kegan; Canto-Sperber and Dupuy, 2001; Perrenoud, 2001; Haste, 2001; Callieri, 2001). Reflectivity is retained as an important transversal feature of key competencies (Rychen, 2003).

6. The three-fold categorization – a conceptual basis for key competencies

In addition to the conceptualization of reflectivity as the required competence level, the demand–oriented approach to competencies resulted in the definition of three broad categories of key competencies – *interacting in socially heterogeneous groups*, *acting autonomously*, and *using tools interactively*.

As an interdependent ensemble, these broad categories form the basis for identifying and mapping competencies that are important in all relevant social fields. This categorization allowed for the extrapolation of nine key competencies, the result of an analysis of the various lists of key competencies presented in the country reports and expert papers in light of the definitional, conceptual, and normative criteria established in DeSeCo.

The three categories of key competencies and the key competencies identified within them have been constructed at a sufficiently general or abstract level to be valid across contexts and nations.

The three-fold categorization has been presented in detail in DeSeCo's final report (Rychen, 2003, p. 85-107). What follows is a brief summary.

6.1 Interacting in socially heterogeneous groups

In this category the focus is on the interaction with the "different other". Individuals need to learn how to interact and coexist with people who do not necessarily speak the same language (literally or metaphorically) or share the same culture, history, or socio-economic background and how to deal with differences and contradiction.

The ability to *relate well to others*, to *cooperate*, and to *manage and resolve conflicts* are particularly relevant in pluralistic, multicultural societies. These competencies include many of the features associated with terms such as "social competencies," "social skills" or "soft skills".

6.2 Acting autonomously

The construct "acting autonomously" concerns the empowerment of individuals to develop a personal identity and to exercise relative autonomy in the sense of deciding, choosing, and playing an active, reflective, and responsible part in a given context.

The ability to *act within the big picture* or the larger context; to f*orm and conduct life plans and personal projects*; and the ability to *defend and assert one's rights, interests, limits, and needs* are critical competencies for participating effectively in different spheres of life – in the workplace, in one's personal and family life, and in civil and political life.

6.3 Using tools interactively

"Using tools interactively", the third category of key competencies, responds, in particular, to the social and professional demands of the global economy and the "information society", which require mastery of socio-cultural tools such as language, information, and knowledge, as well as physical tools such as computers.

Using tools *interactively* does not simply mean having the tool and the technical skills required to use a tool. Rather, it assumes a familiarity with the tool itself and an understanding of how the tool changes the way one can interact with the world and how it can be used to accomplish broader goals. The three key competencies identified in this category are the ability to *use* - interactively – *language, symbols, and text*; *knowledge and information*; and *technology*.

6.4 Key competencies are relevant as constellations

The hypothesis is that for individuals to meet social and economic demands and goals in a successful way, they need to be able to mobilize in a particular context these kinds of competencies.

The notion of *constellation* has been proposed to represent the interrelated character of key competencies and their contextual specificities. The underlying assumption is that meeting any demand or objective will require constellations of key competencies. These constellations vary with the respective context or situation in which they are applied. The specificities and relative importance attributed to key competencies within a constellation may be influenced, for instance, by cultural norms, technological access, social and power relations and the proficiency level of the individual with respect to the enabling tool set.

7. Relevance of DeSeCo for ALL and future assessments

DeSeCo outlines a path for a common understanding of key competencies and their many-layered reference points (normative and conceptual). It offers a comprehensive frame of reference for advancing research and development activities in a coordinated and collaborative fashion to broaden the range of competencies in future assessments and for planning and implementing a coherent, long-term strategy for assessments and indicators of key competencies among young people and adults.

7.1 Towards a common, coherent international discourse on competence and skill development

Terms such as key or core competencies, and life skills or basic skills enjoy much popularity in social sciences and education policy. Often they are used interchangeably or in a vague sense. By defining explicitly the meaning and nature of competence and key competence, DeSeCo provides solid theory-grounded foundations for a common understanding of desired education and learning outcomes in terms of competencies.

The skills and abilities subsumed under the notion of "life skills" used in the ALL framework are an integral part of the psychosocial prerequisites needed to cope with important demands in life and, thus, are consistent with DeSeCo's conceptualization of key competencies. Yet the notion of "life skills" as such seems rather problematic lacking rigor and consistency in public discourse and sometimes also in specialized literature.

DeSeCo's framework, including the concept of key competence, thus offers a useful basis for situating the results from ALL in a broad, internationally consolidated, normative and conceptual context. In turn, by providing relevant empirical evidence on how investing in key competencies can benefit to both individuals and societies, ALL underpins empirically, for instance, the policy discourse of OECD's education ministers on "Investing in competencies for all" (OECD, 2001).

7.2 Situating assessment frameworks and empirical results in a broader conceptual context

Assessments, in particular international assessments, have important policy implications. Situating current domain-specific frameworks and measures of education and learning outcomes in a general frame of reference such as the one constructed in DeSeCo (see figure 1) is a way to recognize the value, but also the limitations, of large-scale studies such as TIMSS, IALS, PISA, and ALL. It allows, when interpreting the results and formulating policy strategy, to take into account the complexity of the topic and to explain the results from different perspectives. The analytic power will increase by addressing not only the competencies or the components as such, but also considering the various factors behind the definition of key competencies: the characteristics of the demands, the linkage to desired outcomes in terms of a successful life and a well-functioning society, and socio-economic and cultural factors.

Also, an overarching frame of reference allows to make more explicit what is actually being measured and what is not and thus counters the risk of over-interpreting the data and of focusing exclusively on competencies for which reliable measures currently exist. It is important not to detract attention, and resources, for education and training from competencies and cognitive skills that are critical in professional and everyday life, but for which to date no adequate methodology for large-scale assessments exist.

For instance, the ALL consortium invested much in the construction of a framework for teamwork (Baker, Horvath, Campion, Offermann, and Salas, 1999) and the development and testing of internationally comparable measures. This effort resulted in the recognition that measuring teamwork reliably would imply observation of individuals interacting in teams, and thus calling for new assessment methodologies (see Murray, 2003, p. 148-149). Thus, the fact that to date, at the international level, it is not possible to provide reliable measures of teamwork does not mean that this key competence is not important in different contexts. In fact, there is a broad consensus on the importance of the ability to cooperate, or work together for common purposes, as many of today's demands cannot be met by one individual alone, but requires individuals to join and function in groups. DeSeCo underpins the importance of key competencies in this domain.

7.3 ALL provides relevant information for the category of using tools

DeSeCo's findings confirm that what IALS and ALL set out to measure – prose literacy, document literacy, numeracy, the analytic reasoning dimension of problem solving, teamwork, information and communication technology literacy and practical intelligence – and what was eventually reliably measured – reading literacy, numeracy, and analytical reasoning - capture critical aspects of key competencies. DeSeCo confirms their theoretical relevance to the category of "using tools interactively" and at the same time, places them in a larger context.

Reading literacy and numeracy can be understood as specificities of the key competence "the ability use language, symbols, and texts interactively" belonging to the category "using tools interactively". DeSeCo emphasizes the need to look beyond the technical skills

It is in the category of using tools that current international assessments can provide empirical evidence on the salience of these key competencies for personal, social, and economic well-being. ALL can provide estimates of the demand for this type of competencies, of their economic, social and educational relevance, and of the implications of the various levels of proficiency for different life domains.

Yet, most of the efforts have been devoted to measuring cognitive skills implied by using a tool, an area for which theory provides a basis for measurement. Other psychosocial prerequisites that need to be mobilized for competent performance in a particular context such as ethical and motivational aspects have not been addressed until now. The PISA Science framework provides promising approaches in this respect.

In addition, future assessments need to address complexity, i.e. assess competencies at the level required by social and economic demands (see notion of reflectivity). For instance, better measures need to be developed on the capacity of individuals to analyze and recognize patterns, to establish analogies between experienced situations and new ones, to discriminate between relevant and irrelevant features, to think and act in a more integrated way, to evaluate actions in light of personal and social goals.

Another important dimension is criticality in terms of the relevance of the competencies to achieving desired personal, social and economic outcomes (e.g. Gilomen, 2003).

Both, complexity and criticality are dimensions that reflect critical features of key competencies.

And finally, although reading literacy and numeracy are important key competencies for young people and adults, they are obviously not sufficient for coping with the manifold demands associated with securing an overall successful life and a well-functioning society. DeSeCo's three-fold categorization offers a basis for establishing priorities about which new competencies should and could be included in future international assessments, and thus guide the systematic extension of future assessment instruments toward capturing a wider range of key competencies. Considerable research and development activities will be necessary to operationalise key competencies related to acting autonomously and interacting in socially heterogeneous groups and to develop internationally comparable measures.

8. A bridge between student and adult competence assessments

Assessments of competencies of young people alone cannot preview the full picture of key competencies of adults. DeSeCo provides theoretical foundations for the development of a coherent long-term program for the assessment of key competencies that includes the whole population (OECD, 2002).

First, the identified key competencies, including reading literacy, mathematical competence, and ICT competence, are relevant throughout the life course in multiple domains (e.g. Murnane and Levy, 2001).

Second, the demands on individuals can be expected to change throughout their adult lives as a result of transformations in technology and social and economic structures (e.g. Haste, 2001). Although more research is needed to conceptualize the interrelationship between competence development and contextual changes, the discussion so far points to the importance of lifelong learning.

Third, developmental psychology (e.g. Kegan, 2001) shows that competence development does not end at adolescence but continues into adulthood. The competence level underlying key competencies (i.e. reflectivity) that would equip individuals to enact successfully the competencies considered necessary for adults in today's societies develop only gradually throughout adulthood. The competence level required by many social and economic demands is reached when, for instance, adults can think for themselves, make their own judgments, and thus act in a reflective and responsible manner.

This understanding has important implications for both education and assessment. The underlying evolutionary model of human development provides a theoretical foundation for the purpose of adult education. Furthermore, it provides a compelling rationale for the importance to assess the competencies of individuals throughout life, and thus to design a coherent overall assessment strategy that spans youth and adults, as already outlined in DeSeCo's Strategy Paper (OECD, 2002).

DeSeCo's overarching frame of reference sets a conceptual context for international assessments such as PISA, IALS, and ALL, and can serves as a bridge between those involved in school-based assessment of key competencies and those assessing key competencies in adults.

9. Challenges for future assessments

There are several aspects of key competencies that will pose important challenges in any effort to assess and measure them. For instance, the notion that key competencies function as constellations raises questions about current usage of assessments for benchmarking based on single measures. As assessments are developed in the future, it will be particularly important to recognize that key competencies do function together as constellations and to identify the strength and direction of any dependencies that exist. In practical terms this implies assessment designs that explicitly allow one to estimate the inter-skill covariance matrix.

As pointed out earlier, developing measures of complexity and criticality are important to reflect critical features of key competencies. In particular, further research and development work is needed to explore conceptually and empirically the assumed linkages between key competencies and desired outcomes at the micro and macro level. The relation of competence constellations to specific social, economic, and political contexts could provide an innovative track for the interpretation of international assessment results.

To do justice to the demand- and context-oriented nature of key competencies and to measure them validly and reliably necessitates a very open approach to assessment methodologies (e.g. Oates, 2003) and the exploration of more adaptive and interactive assessment techniques.

To further advance assessment of education and learning outcomes at the international level, sustained and coordinated research and development work is critical. The international and interdisciplinary collaboration in this domain over several years was beneficial, and should be maintained. As the process continues, the theory-based framework defined in DeSeCo will need to be revisited and refined through confrontation with empirical data and continued interdisciplinary research.

It is important to build future assessment on existing studies, expertise and knowledge, thus not to reinvent the wheel. At the same time, many of the challenges related to education and learning outcomes and their assessment call for new creative and proactive approaches. DeSeCo and ALL have established networks of researchers that can contribute – from different perspectives - to continued research on key competencies and the educational, social, and economic factors that contribute to improve the education and training and to enhanced returns on investments in competencies in terms of personal, economic, and social well-being.

References

Bourdieu, P. (1977). “Cultural reproduction and social reproduction.” In J. Karabel and A.H. Halsey, eds., *Power and ideology in education*. New York, NY: Oxford University Press.

Callieri, C. (2001). The knowledge economy: A business perspective. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 228–231). Göttingen, Germany: Hogrefe and Huber.

Canto-Sperber, M., and Dupuy, J.P. (2001). Competencies for the good life and the good society. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 67–92). Göttingen, Germany: Hogrefe and Huber.

Coleman, J.S. (1988). “Social capital in the creation of human capital.” *American Journal of Sociology*, 94 (suppl.), 95-120.

DeSeCo Project Web site. (2002). Retrieved October 28, 2003, from http://www.deseco.admin.ch

Gilomen, H. (2003). Desired outcomes: A successful life and a well-functioning society. In D.S. Rychen and L.H. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society* (pp.109–134). Göttingen, Germany: Hogrefe and Huber.

Gonczi, A. (2003). Teaching and learning of the key competencies. In D.S. Rychen, L.H. Salganik, and M.E. McLaughlin (Eds.), *Contributions to the second DeSeCo symposium* (pp.119–131). Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Haste, H. (2001). Ambiguity, autonomy, and agency: Psychological challenges to new competence. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 93–120). Göttingen, Germany: Hogrefe and Huber.

Kegan, R. (2001). Competencies as working epistemologies: Ways we want adults to know. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 192–204). Göttingen, Germany: Hogrefe and Huber.

Levy F., and Murnane, R.J. (2001). Key competencies critical to economic success. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 151–173). Göttingen, Germany: Hogrefe and Huber.

Murray, T.S. (2003). Reflections on future international assessments. In D.S. Rychen and L.H. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society* (pp. 135–160). Göttingen, Germany: Hogrefe and Huber.

Oates, T. (2003). Key skills/key competencies: Avoiding the pitfalls of current initiatives. In D.S. Rychen, L.H. Salganik, and M.E. McLaughlin (Eds.), *Contributions to the second DeSeCo symposium* (pp.171–193). Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Organisation for Economic Co-operation and Development. (2002). Definition and selection of competencies (DeSeCo): Theoretical and conceptual foundations: Strategy paper. Retrieved October 28, 2003, from http://www.statistik.admin.ch/stat_ch/ber15/deseco/deseco_strategy_paper_final.pdf

Perrenoud, P. (2001). The key to social fields: Competencies of an autonomous actor. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 121–149). Göttingen, Germany: Hogrefe and Huber.

Rychen, D.S. (2003). Key competencies: Meeting important challenges in life. In D.S. Rychen and L.H. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society* (pp. 63–108) . Göttingen, Germany: Hogrefe and Huber.

Rychen, D.S., and Salganik, L.H. (Eds.). (2001). *Defining and selecting key competencies*. Göttingen, Germany: Hogrefe and Huber.

Rychen, D.S., and Salganik, L.H. (Eds.). (2003). *Key competencies for a successful life and a well-functioning society*. Göttingen, Germany: Hogrefe and Huber.

Rychen, D.S., Salganik, L.H., and McLaughlin, M.E. (Eds.). (2003). *Contributions to the second DeSeCo symposium*. Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Salganik, L.H, Rychen, D.S., Moser, U., and Konstant, J. (1999). Projects on competencies in the OECD context: Analysis of theoretical and conceptual foundations. Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Schleicher, A. (2003). Developing a long-term strategy for international assessments. In D.S. Rychen and L.H. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society* (pp.161–180). Göttingen, Germany: Hogrefe and Huber.

Trier, U.P. (2003). Twelve countries contributing to DeSeCo: A summary report. In D.S. Rychen, L.H. Salganik, and M.E. McLaughlin (Eds.), *Contributions to the second DeSeCo symposium* (pp. 27–64). Neuchâtel, Switzerland: Swiss Federal Statistical Office.

Weinert, F.E. (2001). Concept of competence: A conceptual clarification. In D.S. Rychen and L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Göttingen, Germany: Hogrefe and Huber.

Chapter 3

Moving Towards Measurement: the Overarching Conceptual Framework for the ALL Study

Marilyn R. Binkley *National Center for Education Statistics, U.S.*
Robert Sternberg *Yale University, U.S.*
Stan Jones *Statistics Canada*
David Nohara *Albany, NY, U.S.*
T. Scott Murray *Statistics Canada*
Yvan Clermont *Statistics Canada*

Table of Contents

Table of Contents

1. Introduction

DeSeCo's overarching frame of reference provides a broad conceptual context for the ALL survey development but is not sufficiently detailed to guide the development and validation of reliable, comparable and interpretable approaches to measurement. For this reason the ALL study design team chose to review two core areas—employability skills derived from job and task analysis and psychological theories of intelligence—as a means to move towards measurement. This section sets out how, at least in theory, a framework based upon these two areas can be used to develop "life skills profiles" of people and of job requirements. It concludes by proposing an assessment of a subset of the possible array of life skills identified in the resultant overarching framework.

By definition, moving towards standardized, international comparative assessment implies reducing the range of domains to be considered for development. To achieve this goal the ALL team imposed a series of criteria on candidate skill domains.

First, only those skill domains for which scientifically convincing theory and an established literature existed were considered. Thus, some intuitively appealing skill domains were eliminated for lack of extant scientific underpinnings.

Second, selected domains had to exhibit an accompanying tradition of measurement upon which the ALL study might draw.

Thirdly, current approaches to measurement had to be sufficiently compact to provide valid, reliable and comparable measures within the natural limits of a household survey of very heterogeneous adult populations.

Fourthly, there had to be evidence that the skills in question could be both learnt and, therefore, taught, at least to conventional levels required by everyday life. Thus, skills dependent solely innate abilities were excluded. Similarly, no effort was made to assess the extreme positive range of skill where skill and art merge and where tasks become dependent on specialized lexicons and bodies of knowledge.

Finally, evidence had to exist that the domains to be measured had a direct impact on the social, health, educational or economic life chances of individuals.

In this manner the Adult Literacy and Life skills survey project supports the development and use of the life skills framework and the concept of life skills profiles.

As noted above the work of DeSeCo has expanded and enhanced the theoretical coherence of the life skills framework. In turn, ALL will provide empirical evidence to support and confront this theory building. Given this context, the goal is not to pen the "last word" on life skills, but rather to advance science in this area. The many different efforts to identify life skills have resulted in a large variety of classification systems, each with its own unique nomenclature. As the ability to assess a wider range of skills and abilities increases, it is important to begin to synthesize this continually expanding body of knowledge into an overarching framework to help guide assessment efforts and to build bridges between the various societal actors interested in skills and competence through the provision of valid, reliable, comparable and interpretable data on skills denominated in a common nomenclature.

2. What are Life Skills?

In identifying life skills, it is useful to define what they should represent. As a start, one can be clear about what they *do not* represent. For example, many factors can contribute to one's success in life, but not all of them can be considered "skills." People often attribute their success to such factors as luck, socioeconomic status, physical and social surroundings, fate, or divine intervention. While we do not deny the importance of any of these factors, they are well beyond the scope of ALL. Furthermore, although skills and abilities related to strength, fitness, and physical dexterity have traditionally been important to success in life, ALL chose to exclude any explicit treatment of physical abilities.

It is also important to emphasize that life skills must be connected to success in life. There are many skills, talents, and abilities that do not meet this criterion, even though they may involve sophisticated intellectual processes. This means that not all academic abilities are necessarily life skills, nor are all life skills are likely to be taught in school. This criterion also means that one must recognize that these skills will not be the same—or will not be valued equally—in even a limited range of cultural settings. For instance, one expects that cross-cultural differences in life skills may echo the research on the concept of intelligence. As Sternberg and Kaufman (1998) point out in their review of related literature, at the extreme Western cultures tend to emphasize "technological intelligence" (Mundy-Castle, 1974), generalization or going beyond the information given (Connolly and Bruner, 1974; Goodnow, 1976), speed (Sternberg, 1985a), minimal moves to a solution (Newell and Simon, 1972), and creative thinking (Goodnow, 1976). In Eastern cultures, by contrast, Buddhist and Hindu philosophies stress waking up, noticing, recognizing, understanding, and comprehending, in addition to determination, mental effort, and feelings (Das 1994). African conceptions of intelligence focus on skills that help facilitate and maintain harmonious and stable inter-group relations (Ruzgis and Grigorenko, 1994). But, even in the more limited range of countries included in OECD, variation in how skills are valued is expected.

It is important to note that such variation does not preclude measurement of the underlying skill. Provided that the assessment embodies sufficiently robust theory, that the assessment design affords good coverage of the intended content domain and that the statistical techniques employed to summarize proficiency compensate for the various sources of "missing" data then assessments such as ALL can provide valid, reliable, comparable and interpretable profiles of skill. One of the explicit goals of ALL is, in fact, to explore the variation in how different economies value the skills assessed, how these differences influence the social distribution of economic, social, educational and health outcomes observed at the individual level and at the macro level and how valuations are amplified or attenuated by relative conditions of supply and demand.

Because of this natural variation, the goal in developing a set of life skills is not to establish a single set of benchmarks for people to use to evaluate their successes in life. Instead, one hopes to develop a framework comprised of skills that may have varying importance for different individuals or in different societies but which, when looked at as a whole, accommodate definitions of success used by most individuals and by most societies.

Finally, the definition of life skills should address how they are used. The most common way—and the way that appears in conventional definitions of intelligence—is through *adaptation* to the environment (Sternberg and Detterman, 1986). For example, people must adapt to workplace environments and to new responsibilities as their family lives change. Even people involved in creative endeavours, such as authors, artists, and entertainers, must take into account the tastes of their audiences, markets, or clients, as well as changes in the available technologies of production. Individuals can, however, use life skills to *shape* their environments, such as when a worker modifies a piece of machinery or a production process to increase comfort or efficiency. When neither adaptation nor shaping leads to a successful interaction with the environment, individuals can use life skills to *select* a new environment, such as when a person decides to change workplaces, move to a new location, or become friends with a new group of people.

Therefore, for the purposes of this study, we define life skills as:

> *skills or abilities individuals need in order to achieve success in life, within the context of their socio-cultural milieu, through adaptation to, shaping of, and selection of environments.*

The following sections examine various theories and models related to life skills to see if there is consensus or convergence on particular skills that fit this definition.

3. Deriving Life Skills from lists of workplace skills

The existing body of work on skills necessary for employment success is clearly relevant for our purposes. This perspective has recently received increased attention through the release of several documents setting out lists of such skills (Jones, 1996). These studies and reports cite a need to identify generalizable skills and abilities necessary to better prepare people for success in an ever-changing economy. In so doing, they call attention—sometimes explicitly and sometimes implicitly—to the emerging belief that traditional notions of "basic skills" are not sufficient for success in the workplace. Support for this belief can be found, for example, in the literature on workplace literacy, in which researchers have reported that common school-based notions of reading, writing, and arithmetic are not sufficient for the tasks that adults perform (Sticht, 1978; Mikulecky, 1982; Daggett, 1991). Other researchers looking at job performance found that even in a broadened sense, these three basic skills were not sufficient, and that other skills were also needed (Carnevale et al., 1988). Consequently, new terms and conceptions of basic skills began to emerge.

The term "basic skills" evolved into "employability skills" because these skills were almost always discussed in the context of the transition from school to work or the transition of the unemployed into employment. Although this term is sometimes limited only to those skills necessary for job entry, it usually covers the skills thought necessary to retain jobs to adapt to changes in the organization of work and technologies of production and to secure advancement, such as those that relate to attitude and work habits. Other terms found in the research include "enabling skills," "generic skills," "core skills," "key competencies," " essential skills," and "necessary skills." These different terms would seem to have slightly different implications, but they were often chosen to meet specific local circumstances and preferences, and, thus, are not related in any systematic way to differences in the way these skills were conceptualized. Despite the strong labour-market orientation of these terms and their sources, many of these documents either directly state or imply their relevance to life in general, making them candidates for "life skills."

4. Synthesizing the skills lists

Over the past ten years, a large number of education- and labour-related organizations have undertaken projects to identify employability skills. These include national, state, and provincial government agencies, school districts, and public and private research institutions. Due to this large number, a choice was made to examine a set of these documents that were believed to be representative of the larger group in terms of methodology and findings. These documents include:

- People and Skills in the New Global Economy (Premier's Council, 1990)
- Putting General Education to Work: The Key Competencies Report (Committee to Advise the Australian Education Council, 1993)
- Michigan Employability Skills Profile (Pestillo and Yokich, 1988)
- Conference Board Employability Skills Profile (McLaughlin, 1992)
- *Workplace Basics: The Skills Employers Want* (Carnevale, Gainer, and Meltzer, 1988) (developed for the American Society for Training and Development)
- *Work Keys* (American College Testing, 1997)
- *What Work Requires of Schools* (U.S. Department of Labor, 1991) (a report of the Secretary's Commission on Achieving Necessary Skills (SCANS)
- *National Council on Vocational Qualifications Core Skills* (Oates, 1992) (a system developed for use in the United Kingdom)
- *Essential Skills Research Project* (http://www.hrdc-drhc.gc.ca/hrib/hrp-prh/skills/essentie.html) (a project for Human Resources Development Canada)

The methodology used by these and most of the other studies is generally to start from a broad definition (e.g., "a skill applicable to a wide range of jobs and contexts") and then survey or observe workers, supervisors, and experts to determine what skills are common. Those that pass some test of "frequent enough" across occupations qualify for inclusion in the final list of skills. Aside from terminology, the main differences among the final products have been level of detail and structure. Some simply list a skill (e.g., "reading") while others provide a fuller description of that skill with examples of its application to various situations. Some of the lists provide a single set of skills, but many attempt some sort of categorization or hierarchy. For example, the *Workplace Know-How* from the U.S. Department of Labor's SCANS designates two categories: *Foundation Skills*, which include basic academic skills, thinking skills, and personal qualities; and *Workplace Competencies*, which include the ability to use resources, interpersonal skills, information, systems, and technology (Secretary's Commission on Achieving Necessary Skills, 1991). The Premier's Council Skills Triangle (1990) uses a hierarchy of transferability and generalizability: at the base are *basic* skills, which support *workplace* skills, and these in turn support *job specific* skills. The former two sets of skills are broadly transferable, while the latter set is not.

Despite these differences, a comparison across these lists reveals that they almost always cover the same range of concepts. It should then be possible to distil from them a limited set of categories into which most of the skills listed by any of the nine documents would fit. In some cases, a given list might not address all of the categories in the model, but in as few cases as possible should they contain a skill that does not fit within any of the specified categories.

The table that follows (see Figure 1) presents an attempt to create such a model. Six skill areas were chosen, some with sub-areas, as follows:

- Communication
  - Speaking
  - Listening
  - Reading
  - Writing
- Mathematical
- Problem Solving
- Intrapersonal
  - Motivation
  - Metacognition
- Interpersonal
  - Teamwork
  - Leadership
- Technology

The first column in the table lists these skills. The other columns contain the comparable skills from each of the nine studies/reports that were reviewed, demonstrating how each category is addressed by a given skills list. Skills that did not seem to fit any category were placed in the final row of the table.

Trying to fit a diverse set of models into a framework requires re-arranging the original lists. In addition, although the original terminology from the nine studies has been maintained, placing the terms into the ALL framework involves varying degrees of interpretation. To allow for comparisons to the original sources, the skills lists from the nine documents have been reproduced in Appendix A.

Figure 1
Comparison of skills lists, part I

Skill category	Skills triangle from the premier's council	Key competencies report	Michigan employability skills profile
Communication			
Speaking	Communications	Communicating ideas and information	Speak in the language in which business is conducted
Listening	Communications		Follow written and oral instructions Follow written and oral directions
Reading	Reading and writing Communications		Read and understand written materials Understand charts and graphs Follow written and oral instructions Follow written and oral directions
Writing	Reading and writing Communications	Communicating ideas and information	Write in the language in which business is conducted
Mathematical	Mathematics	Using mathematical ideas and techniques	Understand basic math Use mathematics to solve problems
Problem solving	Analytic problem solving	Solving problems	Use scientific method to solve problems
Intrapersonal			
Motivation			Attend school/work daily and on time Meet school/work deadlines Work without supervision
Metacognition	Ability to learn		Learn new skills
Interpersonal			
Teamwork	Workplace interpersonal	Working with others and in teams	Actively participate in a group Know the group's rules and values Listen to other group members Express ideas to other group members Be sensitive to the group members' ideas and views Be willing to compromise if necessary to best accomplish the goal Work in changing settings and with people of different backgrounds
Leadership			Be a leader to compromise if necessary to best accomplish the goal
Technology	Generic technical	Using technology	Use tools and equipment
Not included	Firm- and job-specific skills Motor skills	Planning and organizing activities	Use research and library skills Use specialized knowledge and skills to get a job done Develop career plans Know personal strengths and weaknesses Demonstrate self-control Pay attention to details Identify and suggest new ways to get the job done

Figure 1
Comparison of skills lists, part II

	Conference board employability skills profile	Workplace basics: The skills employers want	Work keys specifications	SCANS competencies and foundation skills
Communication				
Speaking	Communicate	Listening and oral communication		Basic skills (reading, writing, arithmetic and mathematics, speaking, and listening)
Listening	Communicate	Listening and oral communication	Listening	Basic skills
Reading	Communicate	Reading	Reading for information Locating Information	Basic skills
Writing	Communicate	Writing	Writing	Basic skills
Mathematical	Think	Computation	Applied mathematics	Basic skills
Problem solving	Think	Creative thinking/problem-solving		Thinking skills
Intrapersonal				
Motivation	Positive Attitudes and Behaviours, Responsibility, Adaptability	Self-esteem/goal-setting-motivation/ personal and career development		Personal qualities
Metacognition	Think, Learn for Life	Learning to learn		Thinking skills
Interpersonal				
Teamwork	Work with Others	Interpersonal/negotiation/teamwork	Task Skills Relationship Skills Communication Skills	Use interpersonal skills
Leadership		Organizational effectiveness/ leadership		
Technology	Think		Analyzing the problem Identifying relevant aspects of the problem Understanding the data Choosing appropriate technology Solving the problem Evaluating alternative solutions Extrapolating to novel situations	Use technology
Not included				Use resources Use information Use systems

Figure 1
Comparison of skills lists, part III

	National council on vocational qualifications core skills	Essential skills research project
Communication		
Speaking	Communication	Oral Communication
Listening	Communication	Oral Communication
Reading	Communication	Textual Reading Use of documents
Writing	Communication	Writing
Mathematical	Application of Number	Numeracy skills
Problem solving	Problem-solving	Problem solving, Decision making, Planning/Organizing Job Tasks
Intrapersonal		
Motivation		
Metacognition	Improving own learning and performance	Continuous Learning
Interpersonal		
Teamwork	Working with Others	Working with Others
Leadership		
Technology	Information Technology	Computer Skills
Not included		Psychomotor Skills

Only the Michigan Employability Skills Profile and the SCANS *Workplace Know-How* have a significant number of items that do not have a clear place within the ALL categories. A few, such as “develop career plans” and “pay attention to details” in the Michigan Employability Skills Profile, would seem to be separate skills. Several others, however, appear to involve skills from a combination of the categories in our list. For example, “identify and suggest new ways to get the job done” from the Michigan Employability Skills Profile might incorporate a combination of problem solving, teamwork, and communication. Thus, these skills are not necessarily missing from the ALL model; rather, they might require a restructuring of the categories.

Overall, the ALL skill categories do seem to accommodate the skills identified by the documents that were chosen to be examined and, would do so for most any other skills list found in the employability skill literature. However, it can not be claimed that the combined list is complete. There may very well be other skills or competencies not identified by any of these efforts. If future efforts identify missing elements, then the model should be open to refinement, but the consistency among these studies gives enough confidence to proceed. A more likely shortcoming than omission is that a simple list of skills does not necessarily provide a sufficient degree of depth for a sophisticated understanding of life skills or a useful system of assessments. Therefore, the resulting set of life skill categories was treated as a first step in building an overarching framework of life skills.

5. Deriving Life Skills from Psychological Theories

Efforts to identify formal lists of employability skills are fairly recent developments. In contrast, the effort to describe and measure human intelligence has a history of over 100 years (Sternberg and Kaufman, 1998). The theories developed fall into a variety of paradigms, such as a psychometric paradigm, a cognitive or computational paradigm, a biological paradigm, an epistemological paradigm, an anthropological paradigm, a sociological paradigm, and a systems paradigm (Sternberg, 1990).

Not all of these paradigms are clearly relevant to our discussion. For example, cognitive theories have been applied primarily to tasks used in the laboratories of cognitive psychologists (e.g., Hunt, 1980) and to psychometric tasks (e.g., Sternberg, 1983), but they have not been equally applied to everyday activities. Thus, it is not clear that they meet the criteria of being necessary for success in life. Biological theories are helpful in relating intellectual functioning to the brain (see Matarazzo, 1994) but do not yet carry any implications for how one might go about understanding or assessing life skills. In an adult context, the epistemological paradigm (Piaget, 1972) has proven very useful for evaluating children's sensorimotor, logical, and scientific thinking skills. It has not, however, been shown to be equally useful for analyzing individual differences in adult performance; this is perhaps because the theory was explicitly proposed as a theory of human commonalities, not as a theory of individual differences. Finally, anthropological and sociological theories (e.g., Berry, 1974; Feuerstein, 1980; Greenfield, 1997; Laboratory of Comparative Human Cognition, 1982) point out the necessity of taking cultural and other contextual variables into account but are far from complete as theories of intelligence, much less life skills.

Because of their comprehensiveness and the fact that they are commonly used to discuss practical skills and abilities, the theories of greatest interest here are psychometric theories and systems theories. Even after limiting the scope to these two types of paradigms, one is still left with a number of theories that seem to make very different claims. Despite this apparent impasse, a closer examination reveals that these theories are in fact complementary. An analysis of the two can then lead toward a theory-based concept of a set of life skills, which can then be compared with the set of skills derived from the employability skills literature.

Psychometric theories

Early psychometric theories of intelligence focused on a single general intellectual ability, *G* (Spearman, 1904). Although the concept of *G* is still accepted by many psychometric theorists (see, for example, Jensen, 1998), most modern theories view human abilities as too complex to be captured by a single measure (e.g., Gustafsson, 1988). The large majority of psychometric theorists today accept some kind of hierarchical model with *G* at the top (e.g., Cattell, 1971) or see *G* as existing within a range of academic skills (Sternberg, 1997b). In any case, a life skills model requires concepts of intelligence that can be more clearly and more specifically defined, exemplified, and assessed than this broad and elusive concept.

Representative of the hierarchical theories are Cattell (1971), Carroll (1993) and Horn (1994). They describe hierarchies moving from the specific—e.g., spelling ability and speed of reasoning in the Carroll model—to the general—*G* in the Cattell and Carroll models. While *G* is too general for current purposes, it is not clear whether all or most of the abilities found at the opposite end of the hierarchy (i.e., specific abilities) would fit our definition of life skills as being necessary for success in life. For the current purpose, the most relevant and useful abilities included in these models are found somewhere in the middle.

Two "middle-level" abilities common to most hierarchical models are *crystallized abilities* (also referred to as *Gc*) and *fluid abilities* (also referred to as *Gf*). Carroll describes crystallized abilities as the accumulated knowledge base, including language development, verbal and printed language, comprehension, and lexical knowledge. Similarly, Horn defines crystallized abilities, or acculturation knowledge, as

> "…measured in tasks indicating breadth and depth of the knowledge of concepts and forms of reasoning that have been developed by humans over the course of many centuries and passed on from one generation to the next. *Gc* can be thought of as the intelligence of the culture that is incorporated by individuals through a process of acculturation"(p. 443).

Carroll's description of fluid abilities centers on reasoning abilities, namely, general sequential reasoning, inductive reasoning, and quantitative reasoning. Horn defines fluid abilities as measured in tasks requiring inductive, deductive, conjunctive, and disjunctive reasoning to arrive at an understanding of relations among stimuli, to comprehend implications, and to draw inferences (p. 443).

The ALL framework assumes that these two abilities, crystallized and fluid, are essential to a model of life skills for several reasons. First, it is easy to conceive that acquisition of them in some way or another is necessary for success in life. Second, they appear distinct—an important quality as many of the skills and abilities uncovered in the search for life skills are similar or overlapping. Whereas tests of fluid ability primarily measure the results of current information-processing skills (reasoning), tests of crystallized ability, in contrast, primarily measure the result of previously applied information-processing skills (comprehension). In addition, although measures of the two abilities may show some statistical correlation, the number of studies separating them as distinct factors (Carroll, 1993) is so large that there seems to be little argument for their relative, although not total, independence. Finally, the two abilities show different patterns of growth and decline with age (Horn, 1994). Fluid ability tends to increase from infancy onward until the late 20s or early 30s and then to start a period of usually gradual decline. Crystallized ability tends to increase until rather late in life and only then begins to show some decline.

Other abilities frequently appear in psychometric theories. Most prominent among them is *spatial visualization* (*Gv*), an ability associated with spatial problem-solving (Horn, 1976). While one must keep these other abilities in mind, of the abilities found in psychometric models, crystallized and fluid abilities are the most distinct and widely accepted, and, thus, present the most compelling case for inclusion as distinct components of the ALL life skills framework.

6. Successful intelligence

Even if one were to consider spatial visualization and other less commonly mentioned abilities in addition to the predominant fluid and crystallized abilities, psychometric theories alone do not provide a comprehensive set of intellectual abilities necessary for success in life. Because they have their origins in their attempts to describe and measure different types of thinking, they are less focused on how people relate to their surroundings, which is a primary criterion of life skills. Sternberg (1985, 1997a, 1997b) has attempted to address this shortcoming through his three-part (triarchic) theory of *successful intelligence*. He uses this term to emphasize the importance of intelligence as the abilities needed for life success, thereby distinguishing it from the narrower conception of intelligence that is popular in the psychometric literature.

Sternberg argues that there are three major aspects of successful intelligence—analytical, creative, and practical. Analytical abilities are those abilities used to analyze, evaluate, judge, compare, contrast, and critique. Creative abilities are those abilities used to create, invent, discover, suppose, imagine, and hypothesize. Practical abilities are those abilities used to put into practice, apply, use, and implement knowledge and skills. The three sets of abilities are hypothesized to be statistically relatively independent but to be psychologically intertwined because the same information-processing components underlie all three. What differ are the situational contexts in which the components are used. Analytical abilities tend to be used in situations that are relatively more familiar, creative abilities in situations that are relatively less familiar, and practical abilities in situations that are highly contextualized with respect to the individual's daily life.

It is important to note that the term "creative abilities" does not refer to the high levels of creativity shown by world-famous authors, artists, or scientists. Rather, it is being used in the much more mundane sense of people's abilities to deal flexibly with relatively unfamiliar problems, abilities to cope with relative novelty (Sternberg, 1985). For example, creativity in the sense it is used here is involved when workers on an assembly line formulate a strategy for more quickly moving their parts of products down the line (see Scribner, 1984). These abilities have appeared in only in a minority of psychometric theories of intelligence (e.g., Guilford, 1967).

The theory of successful intelligence adds two important elements to the consensus view that has emerged from the consideration of psychometric theories. First, it emphasizes the importance of the distinction between coping with familiar and with novel tasks and situations. Research suggests that people who are able to cope well with familiar tasks and situations are not those who are always flexible in coping with novel tasks and situations (Sternberg, 1985; Sternberg and Lubart, 1995). But in any job or personal relationship, for that matter, one will be confronted with novel tasks. An assessment of life skills therefore needs a balance of both. This idea is implicit in the distinction between fluid abilities (which can be, but need not be, applied to more novel tasks) and crystallized abilities (which are typically applied to more familiar tasks); the theory of successful intelligence simply makes this point explicit.

The theory of successful intelligence also emphasizes the importance of the processes rather than just the products of intellectual functioning. This emphasis has been implicit in many psychometric theories. For example, although Spearman is most well-known for his structural theory of general ability, he published an entire book on the processes he proposed might underlie *G* (Spearman, 1923). His book detailed three of the so-called "qualitative" processes—apprehension of experience (what is called "encoding" in the theory of successful intelligence), eduction of relations (what is called "inference" in the theory of successful intelligence), and eduction of correlates (what is called “application” in the theory of successful intelligence). Thurstone’s (1938) theory, although not as process oriented, was seen by its originator as a step along the way toward a process theory (Thurstone, 1947). And Guilford’s (1967) theory had processes as one of the three dimensions of the cube that Guilford proposed to represent the structure of intellect.

The degree to which these three abilities are distinct from those proposed by psychometric models—and therefore require treatment as discrete components—can be seen in how well they are measured by conventional psychometric assessments. Analytical abilities, with their close relation to reasoning abilities (fluid abilities), are measured fairly well by conventional assessments. Creative abilities tend to be measured less well by such tests, although the extent to which conventional tests measure creative abilities seems to be proportional to the novelty of the test material. Thus, fluid ability tests that are relatively novel, such as the Raven Progressive Matrices, probably tap into creative abilities, at least more so than do other tests (Sternberg and Lubart, 1995). (In this case, however, it is important to emphasize that what captures creative abilities is not the fact that these are fluid ability tests—which would imply similarities to fluid ability—but their unfamiliar nature.) Practical abilities are measured least well by conventional tests (Sternberg, et al 1995), as evidenced by accounts of people whose ability to develop fairly complex procedures for executing tasks related to their jobs cannot be predicted by their IQs (Sternberg and Kaufman, 1998).

7. Proposed domains for Life Skills derived from psychological theories

For the current purpose, the theory of successful intelligence is best seen as extending along the same dimension of psychometric models. Both describe types of thinking and have some degree of overlap. A review of psychometric models points to the importance of crystallized and fluid abilities. The successful intelligence model begins with analytical abilities, which can be seen as overlapping at least fluid abilities and perhaps crystallized abilities as well. But the successful intelligence model takes us further, addressing people's relationship to the environment through the domains of practical abilities and creative abilities. Thus, as a set of core domains of intelligence with clear relationships to success in life, would include:

- Practical Abilities—Abilities used to practice, apply, use, and implement knowledge and skills. These abilities are highly contextualized with respect to the individual's daily life and involve the management of oneself, management of others, and management of tasks.
- Crystallized Analytical Abilities—Acculturation knowledge. Evident in tasks that show an indication of the breadth and depth of the knowledge of concepts and forms of reasoning that have been developed by humans over the course of many centuries and passed on from one generation to the next. Tests of crystallized abilities primarily measure the result of previously applied information-processing skills.
- Fluid Analytical Abilities—Reasoning abilities, such as sequential, inductive, deductive, and quantitative. Tests of fluid abilities primarily measure the results of current information-processing skills.
- Creative (Coping with Novelty) Abilities—Abilities to create, invent, discover, suppose, imagine, and hypothesize. Characterized by people's abilities to deal flexibly with relatively unfamiliar problems—that is, their abilities to cope with relative novelty.

8. Connecting the findings: A proposed Life Skills framework for the ALL study

In the preceding sections, attempts from two widely different fields to describe skills or abilities that people need to succeed in life were examined. In both cases, a variety of documents were reviewed and frameworks of categories based on convergence around certain skills and abilities were drawn from them. With regard to the documents from which they were drawn, the frameworks are judged to be fairly complete: no commonly accepted skills, abilities, or categories were omitted. It would seem safe to assume, therefore, that they could accommodate most of the theories and models in their respective fields.

It is perhaps no surprise that these two efforts with seemingly similar topics reach such very different conclusions. They do, after all, take very different approaches in examining human abilities. The employability skills models look at the skills required by the many tasks that people face in the workplace and create general categories that are applicable across a variety of situations. The psychological models, on the other hand, look at only the type of thinking in which people commonly engage, with little regard to the context. Neither approach is inherently wrong; both can be considered appropriate given the different interests and perspectives of the fields from which they originate. If the major difference between the two is one of perspective, and the two approaches are indeed talking about a the common concept of life skills, or at least something roughly comparable to it, then an examination of life skills should be enhanced by incorporating both of these perspectives.

The All study designers propose that the relationship between the two models is as follows:

The skills derived from the employability skills literature are the context areas in which the four types of thinking derived from the psychological theories take place; and conversely, in each of the skill categories, people can engage in primarily four types of thinking represented by the four categories derived from the psychological theories. This relationship is represented by the matrix in Figure 2.

According to this proposed relationship, the skills within a skill area can be classified by the type of thinking they involve. Taking the example of mathematical skills, crystallized mathematical skills, such as recalling mathematical facts and formulas, easily come to mind. Fluid mathematical skills might allow a person to solve mathematical problems, whereas practical mathematical skills would enable a person to apply his/her mathematical skills to a situation found on the job. Finally, creative mathematical skills might, for example, allow a person to design a mathematical solution for a seemingly non-mathematical situation. Each of these cases occurs in mathematics; yet there are clear differences among them.

If one looks at the four types of thinking, providing an example of any one of them requires describing a context, such as using creative thinking to develop a new software application, or using fluid thinking to select an appropriate dosage of medicine based on the instructions on the package. The examples might describe a highly specialized situation, such as identifying a faulty part on an airplane engine or writing a computer program to predict seismographic activity. In the vast majority of cases, however, they all can be described, at some level, by a skill found within the skills categories derived from employability skills studies.

Linking the two perspectives in this manner is not simply a convenient way of reconciling two seemingly disparate models. It makes discussions of life skills more focused by adding a greater level of precision to the employability skills categories and providing a means for describing the thinking abilities. For example, given two individuals leading teams, one might perform their duties adequately by employing strategies common within the company, while the other might devise a new strategy more tailored to the project and the team members. Both would be employing teamwork skills, but it would be inaccurate to say both were using the same type of skills. Similarly, it is easy to imagine a person who can calculate the amount of carpet required to cover all the floors in a house but who might not understand a set of instructions explaining how to do so. In this case, is it fair to say that the person has practical abilities, or is it more appropriate to say that he or she has practical mathematics abilities but few practical reading skills?

Describing situations to characterize each of the cells in our matrix forces us to re-examine the meanings of the terms in each model and the relationships of the categories to each other within the same model. In most cases, the differences between the skill categories appear clear and the categories appear fairly discrete (although one might argue that problem solving can take place within a mathematical or technological setting, or that using technology requires communication and mathematical skills). In contrast, as one moves across the domains of thinking skills, they are not necessarily independent of one another. In the example cited earlier regarding leading a project team, the individual devising the new strategy—a creative thinker—may very well base those new strategies on what he or she has learned in business school or commonly used at the company, thus tapping into crystallized thinking. Furthermore, creating a new strategy may require an evaluation of how well different strategies would work, which requires fluid thinking. Without further exploration, it is premature to label these four categories as constituting a hierarchy, but it is important to recognize the possibility that this relationship exists.

It is also premature to provide anything other than tentative examples for each of the cells. In the table on the next page (see Figure 2), examples have been placed in some of the cells. In some cases, they represent commonly occurring tasks and situations that are distinct from others in the same row or column. In other cases, the examples may seem more obscure, contrived, or similar to other examples in the same row or column. It must be recognized that further work in developing examples for each cell may require refinements to the model. And, perhaps more importantly, this ultimately may be a product of the ALL assessments themselves.

Figure 2
Tentative examples of life skills

	Practical abilities (standard applications-one's own life)	**Crystallized analytical** (recall)	**Fluid analytical** (reasoning, information processing)	**Creative abilities** (novel situations)
Communication	——	——	——	——
Speaking	Speaking in different situations	Speaking (diction, clarity, using words correctly)		Speaking to achieve goal
Listening	Listening in variety of situations	Comprehension of spoken words	Understanding and interpretation of meaning	Understanding non-standard speech or type of material
Reading	Reading variety of documents	Reading from text	Understanding and interpretation of meaning	Understanding non-standard language use or material
Writing	Writing various types of documents	Mechanics of writing and grammar		Writing (fiction or non) to achieve particular goal
Mathematical	Applying math to everyday situations	Recall of facts, theorems, and formulas	Solving equations or doing geometric proofs	Designing mathematical solution strategies
Problem solving	Choosing from existing set of solution strategies	Knowledge of standard problem-solving strategies	Deducing solution to problem	Designing solution strategies
Intrapersonal	——	——	——	——
Motivation				
Metacognition				
Interpersonal	——	——	——	——
Teamwork	Participating in team	Comprehension of one's role in group		Taking on new, unfamiliar assignment
Leadership	Applying strategies	Knowledge of common team structures and strategies	Identifying strengths and weaknesses of team members	Designing strategies (based on goals, resources, strengths of team members)
Using technology	Using different software applications	Keyboarding, logging on	Troubleshooting	Creating software for unique situations; devising information search strategies

The following four charts provide a graphical summary of how, ultimately, the ALL study conceives of skill. Figure 3 is meant to convey the dynamic or adaptive dimension of the approach. Individuals are required to confront new demands, whether from changes in technology and life course or because of a conscious desire for change. To confront these changes individuals can draw on a number of possible resolution processes including their own cognitive resources. Different approaches to resolution lead to different outcomes for individuals, and, through aggregation, to the societal level. The relationship between resolution approach and outcome is probabilistic rather than deterministic.

Figure 3
A modified DeSeCo framework for key competences

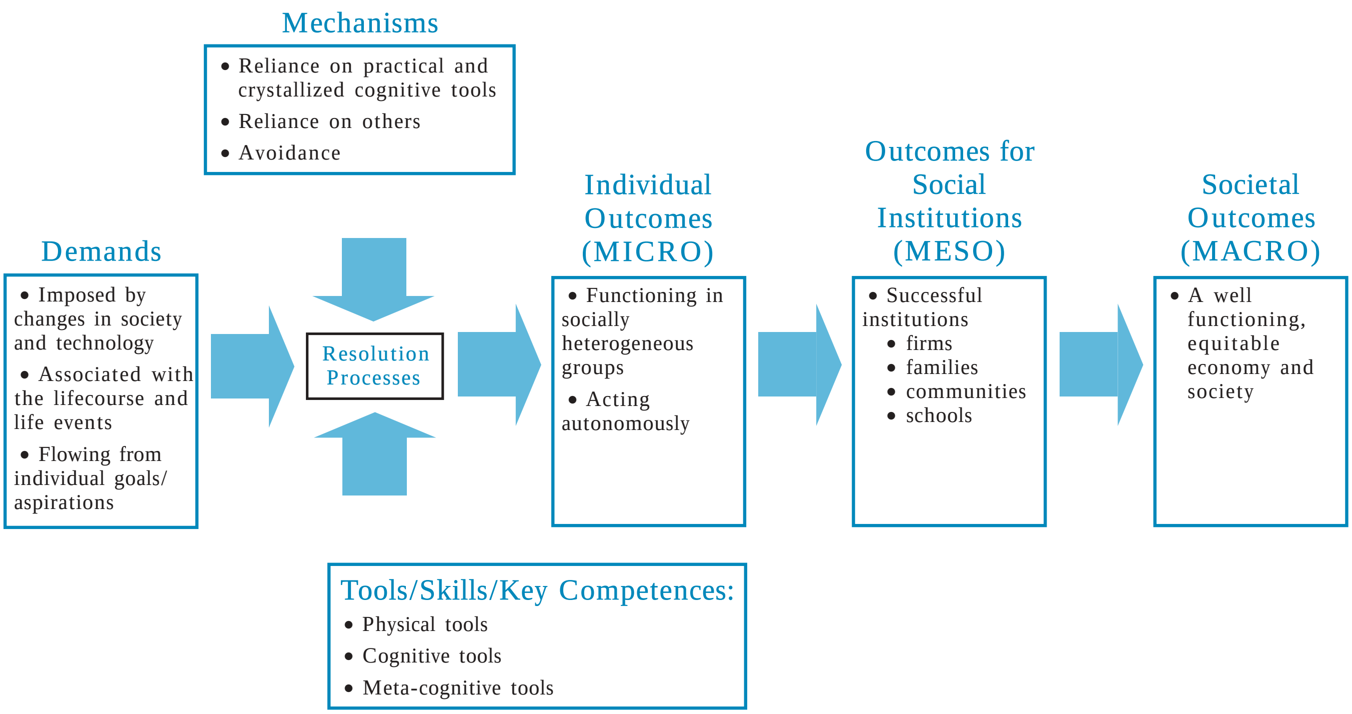


Figure 4 is meant to convey two additional insights. First, it suggests that skill demand can be mapped in multiple life contexts, - work, home, the community, using a standard typology of skill. Second, a hierarchy of skill exists in which the acquisition of skills at higher levels in the skill pyramid depend, again in a probabilistic sense, upon having acquired lower order skills.

Figure 4

Skill supply and demand by context

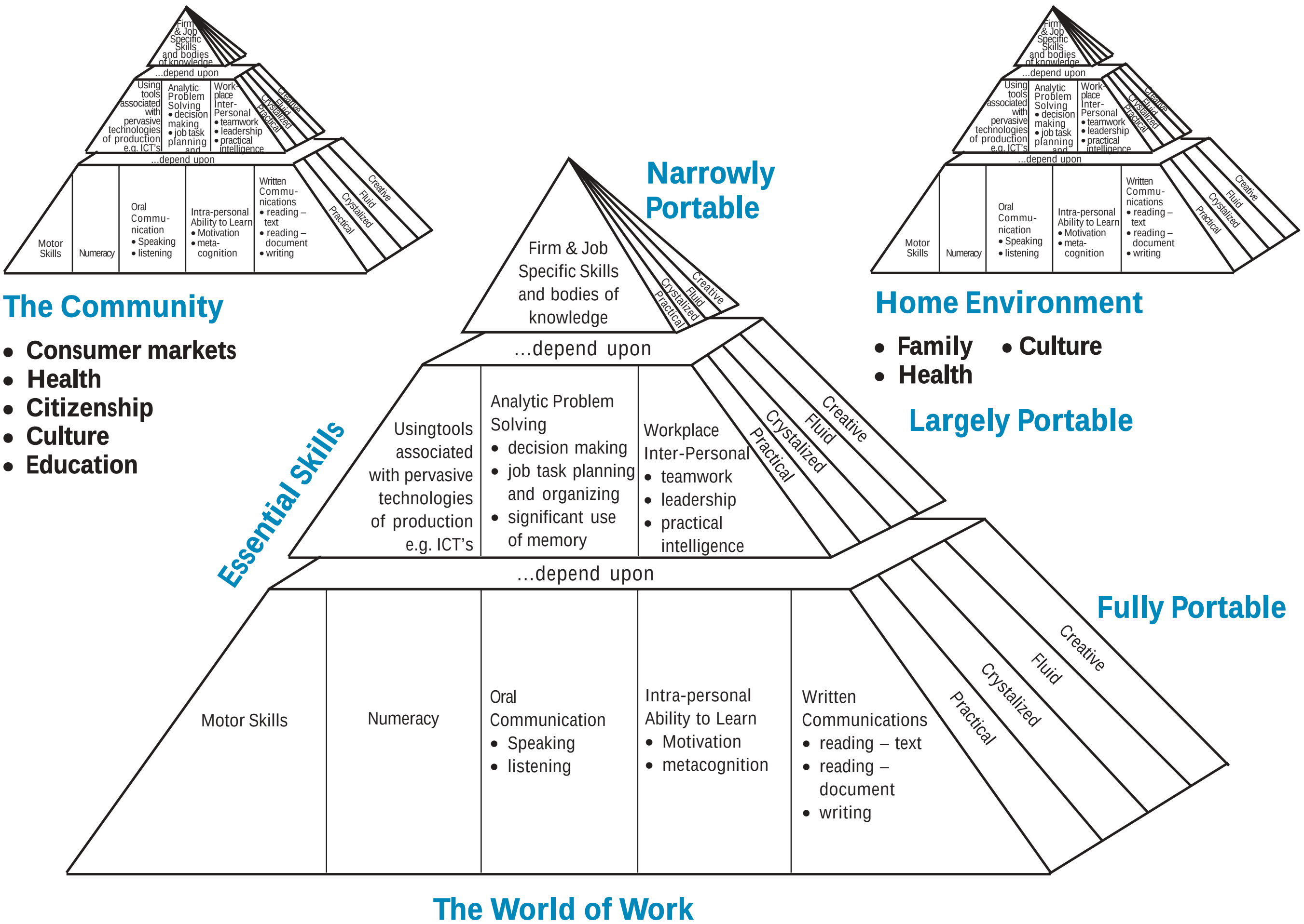
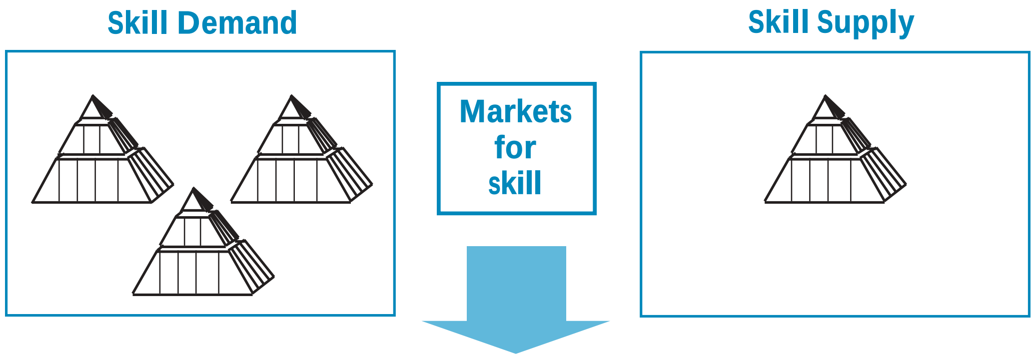


Figure 5 shows how skill supply, profiled in the same skill typology as skill demand, meets demand in multiple markets, the most obvious of which is the labour market. Other less obvious markets include the education market that serves to allocate access to scarce educational resources, health markets that allocate access to scarce health resources including health information, and consumer markets for products and services, many of which are heavily reliant on the printed word and, increasingly upon technology – assisted transactions. Less obvious, the ALL framework thinks of families as markets, or transactional systems, wherein skill demands confronting the family unit are matched with the skills available in the family unit. Skill differences and role specialization among family members can create inequities in the distribution of power along gender lines. Finally political systems can be thought of a market that affords individuals access to power. The works of Feriere (Feriere 1970) and Bordieu (Bordieu, 1977) suggest that access to power in these systems depends critically upon individuals having basic literacy and numeracy skills.

Figure 5

Theoretical framework: A "Markets" model of skill



Outcomes

Context	MICRO (Individuals)	MESO (Social Institutions)	MACRO (Systems)
• Economic			
• Social			
• Educational			
• Health			

A defining characteristic of markets is that they create winners and losers, in this instance based upon individual skill levels. This fact is reflected in Figure 5 as markets influence the distribution of individual outcomes in the economic, social, health and educational domains. These same markets also influence the distribution of outcomes realized by a range of social institutions, such as firms, families, schools and communities, and ultimately the distribution of outcomes observed at the macro level in these same domains. Figure 6 provides examples of outcomes in each domain and level. A less obvious characteristic of markets for skill is that they can be relatively more or less efficient allocation mechanisms. For example, labour markets are thought of as being information poor, a fact that leads to high transaction costs and considerable inefficiency in the employee-employer matching process.

Figure 6

Outcomes flowing from markets for skill

	MICRO (individuals)	MESO (firms, communities, schools, families)	MACRO (economies, societies, regions, special population)
Economic	• employability • wages • reliance on social transfers	• firm profitability • productivity • adaptability of firms and communities • power distributions within families	• overall growth rates • speed of adjustment
Social	• volunteering • community participation		• trust • social capital
Health	• physical health • mental health • mortality • morbidity	• institutional efficiency • insurance costs	• equity • opportunity costs
Educational	• access • persistence to completion • skill level	• inclusion • average • quality	

This final observation allows one to think about how public policy might seek to influence human capital. First, it can serve to increase the supply of skill available to markets, largely by financing learning. Second, it can attempt to influence the demand for skill in each of these markets at the individual level or at the level of social institutions. Finally, it can try to improve the efficiency of markets that match supply and demand for skill. As with any set of market oriented policies care must be taken to ensure that skill supply and demand are roughly in balance over the long term or market distortions and failures will emerge.

9. Applicability of the framework to other theories and perspectives

For the ALL overarching framework to be valid, it should be able to accommodate not only the theories from which it is derived, but also other attempts to describe life skills, types of thinking, employability skills, and similar concepts. It is not necessary that the proposed model agree in either terminology or level of detail with other models and theories, but it does need to be conceptually compatible to them. Many theories regarding skills, abilities, and intelligence sound quite different from the proposed framework. Often, however, the differences arise from the fact that these other theories go beyond identifying skills and abilities to describing how they are acquired, developed, and influenced by family and society. This is the case with most sociological and anthropological theories, for example. Yet, if there are theories with a purpose similar to that proposed that include skills the ALL model does not, or if they contain a structure proposing a different relationship among the elements identified, then the proposed framework will need to be re-examined. Similarly data from the ALL survey itself may suggest modifications to the underlying theory.

One prominent theory of intelligence that has not been examined thus far is the theory of *multiple intelligences* developed by Howard Gardner (Gardner, 1983). It is a psychological theory that identifies eight discrete kinds of intelligence: linguistic, logical-mathematical, spatial, interpersonal, intrapersonal, naturalist, bodily-kinesthetic, and musical. These intelligences have been identified via a retrospective analysis that has allowed for the consideration and addition of new intelligences that can meet established criteria for discreteness. Although it is a psychological theory, it does not always fit along the same dimension as the psychometric theories and the theory of successful intelligence we have considered.

There are certain overlaps with the psychometric theories. One could argue that Gardner's description of linguistic intelligence is very closely related to crystallized abilities and that logical-mathematical intelligence sounds similar to fluid abilities. At the same time, all of the intelligences could also be seen as domains in which intelligent thinking processes—such as those described by psychometric models and the model of successful intelligence—can take place. One can easily conceive of crystallized, fluid, practical, and creative thinking taking place within each of the intelligences. In the case of linguistic intelligence, for example, one can read a short story (crystallized), analyze it (fluid), devise a way to apply the message of it to everyday life (practical), or write an alternative ending to it or write a short story of one's own (creative). Thus, in many ways, these intelligences are more similar in nature to the skills identified from the employability skills literature.

Several of the intelligences, however, go beyond the skills listed in the proposed ALL framework. Whether naturalist intelligence, bodily-kinesthetic intelligence, and musical intelligence are truly intelligences is moot. ALL assumes, however, that their generalized relevance to job performance might be less than that of other skills or intelligences. For example, few jobs outside the field of music require any demonstrated level of musical intelligence. While many people who pursue musical interests find that it enriches their lives, it would be difficult to argue that musical intelligence is necessary for success in life for anything but a small proportion of the population. A similar argument can be made regarding naturalist intelligence.

Overall, comparing the theory of multiple intelligence against the proposed life skills framework reveals that, while the intelligences do not fit neatly only as skills (the left-most column in the matrix) or only as domains of thinking (the top row), the type of abilities they describe are captured in various places throughout the framework. Comparisons with other theories of intelligence or frameworks of employability skills would result in a similar degree of consistency.

10. Envisioning Life Skills "Profiles"

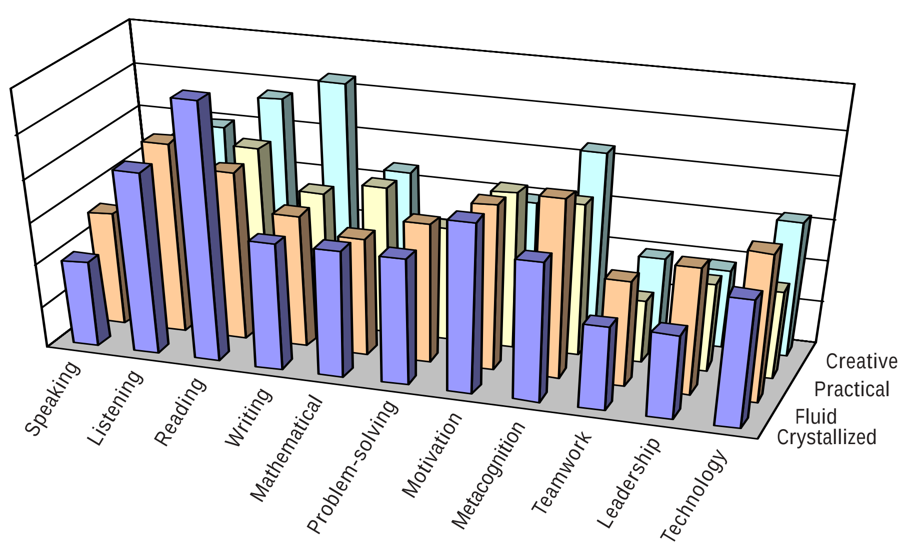
The proposed framework describes those skills that a comprehensive set of life skills contains. If it is valid, it should be relevant across a wide variety of people and contexts. Where differences arise, however, is in the *amount* of each skill a person possesses or a context requires. In other words, while most people need communication skills, some people have and need more than others, such as in the case of a reporter or a counsellor. Similarly, a particular occupation might require all four types of thinking skills, but might rely less on creative than on crystallized abilities. Extending the proposed framework to address skill levels creates a potentially powerful tool for describing people and contexts (e.g., cultures, occupations, and lifestyles).

Figure 7 presents a hypothetical life skills profile for a single life context. It is based on the matrix of the proposed life skills framework, and indicates the relative amounts of each skill the person possesses. In this case, the person possesses fairly strong communication skills and has strong creative abilities in some instances. Rather than focusing on a particular person, the same profile could be used to describe skills required of a particular job. Profiles such as this one could be developed to describe both individuals and occupations and to compare similarities and differences in strengths and emphases.

It is important to remember that these profiles are highly theoretical by nature. The ability to conceive of life skills far exceeds the ability to assess them, and even further exceeds the ability to assess them as a unified set within practical constraints by using comparable methodology and scales. The profiles do, however, serve as useful heuristic devices, not only to help envision the potential applications of a life skills framework, but also to better understand and refine the framework itself.

Figure 7

A Theoretical Life Skills "Profile" for a single life context



11. The third dimension of life skills: Context

Although the ALL framework was derived in large measure from the literature on job and task analysis it is obvious that these same skills are relevant to life contexts other than work. Adults assume multiple roles in everyday life – they are citizens, consumers, parents, caregivers, volunteers, friends and lovers. Each of these roles present adults with problems to solve, be it writing a cheque, reading instructions on a medicine bottle, giving someone instructions, or deciding which clock to buy.

Individuals have the choice of:

- avoiding or delaying such problems
- relying on others to solve their problems, or
- solving the problems themselves.

Those who choose to solve their own problems can do so by calling upon different combinations of skills and by applying different ways of thinking.

Two fundamental assumptions underlying the ALL study are that:

- individuals with low skill levels will be placed at a disadvantage as measured on a number objective social, health, educational and economic outcomes, whereas those with high skills will be advantaged
- demand for the skills measured in ALL are growing rapidly in most OECD member states, a fact that suggests that they will play an increasingly important role in defining outcomes in all life contexts and at all levels.

Rapid changes in the dominant technologies of production and related changes in the organization of work have precipitated rapid changes in the demand for skill.

Increasingly, workers are asked to solve problems themselves using a combination of literacy, numeracy and computer literacy skills, including understanding information on their own benefits responsibilities, expected quality and occupational health and safety.

In the health domain individuals are increasingly expected to manager their own health needs, a task that depends on their health literacy and general literacy levels, as well as their ability to access information on the Internet.

In the consumer domain, buyers are expected to research their purchases to buy goods and services over the telephone and/or over the internet.

Parents are expected to contribute to their children's education, to read to their children and to help with homework and to interact with the school.

Citizens in pluralistic democracies are expected to develop opinions on a range of policy topics and to judge the merit of alternative responses – opinions that are best formed through reading and reflection. Basic numeracy skills are essential to understanding the complexity of the science underlying many policy debates.

Each of the ALL specific domain framework incorporated items drawn from the following life contexts: home and family, health and safety, work, societal, community and citizenship, consumer economics and further learning.

It is important to note that this is not meant to imply that individuals cannot, or do not, manage to solve a broad range of everyday problems without employing the skills assessed in ALL . Clearly they can, and do. Analysis of data from the IALS study suggests, however, that individuals would gain more independence, and would be more successful on average, if average skill levels were increased and the proportion of individuals at level 1 on each scale, the lowest skill level reduced. These effects are sufficiently large to imply large effects at the level of the firm and the community and, ultimately, at the level of entire economies and societies.

12. Measuring Life Skills

The ultimate purpose of the proposed life skills framework is to help guide the assessment efforts of the ALL project. Due to the variety of work being done in the area of assessing life skills and cognitive abilities, the assessment efforts within the ALL project could follow very different paths. The proposed framework provides a unifying direction for assessment efforts by establishing a limited set of life skills to be assessed and by establishing relationships among life skills according to their skill area and the type of thinking involved. Having a vision for what an ideal set of assessments would look like not only ensures consistency in development, but also assists in gauging progress along the way.

The proposed framework requires that any assessment efforts be viewed in terms of two characteristics: the type of skill assessed (e.g., reading, teamwork) and the type of thinking assessed (e.g., crystallized, creative). Doing so makes the development of assessments all the more challenging, as efforts to assess just one characteristic are themselves incomplete. As described earlier, there have been many efforts to identify employability skills, but few of the projects have assessment efforts associated with them. Assessments measuring different types of thinking ability have a considerably longer history, but conventional methods primarily focus upon crystallized and fluid thinking, while assessments of practical and creative abilities are less well developed and not commonly accepted.

Not surprisingly, if we compare existing ALL assessment efforts with the proposed framework, they only address a fairly limited range of life skills.

Several potential domains did not meet the criteria for inclusion set out earlier and were dropped from consideration for the current cycle of assessment. Thus ALL attempted to develop detailed assessment frameworks, test blueprints and performance assessment items in the following seven areas:

> *Prose Literacy (PL)*—focuses on the knowledge and skills needed to understand and use information from texts that contain extended prose organized in a typical paragraph structure found in materials such as editorials, news stories, brochures and pamphlets, manuals, and fiction.
>
> *Document Literacy(DL)*—focuses on the knowledge and skills required to locate and use information in qualitatively different printed materials that contain more abbreviated language and use a variety of structural devices to convey meaning. These include tables, charts, graphs, indices, diagrams, maps, and schematics.
>
> *Numeracy(N)*—addresses the ability to interpret, apply, and communicate mathematical information in commonly encountered situations (adapted from Queensland Department of Education, 1994). Numerate behaviour is observed when people manage a situation or solve a problem in a real context, and involves responding to mathematical information that may be represented in multiple ways; it requires the observation of a range of enabling knowledge and behaviour processes. In the ALL context we refer to numeracy as: the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations.
>
> *Problem Solving(PS)*—involves the ability to apply cognitive processes toward determining a solution when that solution is not immediately obvious to the problem solver. The ALL definition of problem solving takes a "project approach" that focuses on the competencies needed for the regulation of actions in complex tasks.

Teamwork(T)—focuses on the core skill competencies associated with successful and effective teams and teamwork. To this end, three primary competencies required for effective Teamwork are proposed:

Group Decision Making/Planning, Adaptability/Flexibility, and Interpersonal Relations.

A fourth competency, Communication, underlies all three competencies and serves as a bridge between them.

Practical Cognition(PC)—refers to knowledge relevant to problems that are not clearly defined, are personal in nature, relate to everyday experience, have multiple correct solutions, and have multiple methods for deriving solutions. Measurement is achieved through items that present a real-life situation.

Information and Communication Technology Literacy(ICTL)—includes access to computers, self-assessment of computer-related attitude and ability, use of and experience with computers, methods used in the development of computer skills, and use of and experience with related technology.

Figure 8 presents what a cross-mapping of ALL assessments to the framework looks like, placing the initials of each assessment in appropriate cells where assessments took place in the developmental phase.

Figure 8
Current ALL frameworks

	Practical abilities	Crystallized abilities	Fluid abilities	Creative abilities
Communication				
Speaking				
Listening				
Reading	PL, DL	PL, DL	PL, DL	
Writing	PL, DL	PL, DL	PL, DL	
Mathematical	N	N	N	
Problem solving	PC	PS	PS	
Intrapersonal				
Motivation				
Metacognition				
Interpersonal				
Teamwork		TW		
Leadership				
Using technology		ICTL		

Clearly, future assessment will need to expand across a wider range of abilities, creative abilities in particular. Although coping with novelty and flexible thinking are required in some of the assessments, none of the assessments are explicitly designed to tap flexible thinking and the ability to cope with novel kinds of tasks and situations. This is, of course, quite challenging, as these abilities are the most difficult ones to measure and the ones with which psychometricians have the least experience.

In many regards, the ALL assessments are venturing into new territory. Because the entire range of life skills is large, it may not be practical or desirable to address all of the cells in the life skills framework. Priority areas will have to be chosen. However, as efforts to develop assessments evolve, the framework can serve as a valuable method for identifying those priority areas, and for ensuring philosophical consistency across all of the assessments.

13. Conclusion

The framework for life skills presented here is ambitious—it seeks to establish a unifying relationship among theories about skills and intelligence that have widely varying purposes, perspectives, and language. As a result, one might argue that it is incomplete, that it takes liberties with the original theories, or that the relationships proposed are less valid in some instances than in others. It derives strength, however, from the fact that it is based on consensus found within the two types of theories it examines: employability skills and psychological models. Further, it generally accommodates other theories, examples of contextual applications of skills and thinking abilities, and the assessments developed for the ALL project. Future work in the area of life skills may require modifications to the framework but, even in its initial state, it provides an important theoretical foundation for the ALL project. Not only does it address the many different models and theories that have been developed to describe concepts similar to life skills, it distils a comprehensive set of life skills from this convergence. It is intended that this set can accommodate other theories and models, so that future consideration of them does not result in completely new sets of skills and abilities that need to be addressed. The framework also creates a useful means of examining the assessment efforts of the ALL project. By placing them into the framework, it is possible to gauge how "complete" the efforts are and to identify possible areas for future assessment development. Finally, the framework can contribute to the overall vision for the project, focusing both theoretical and practical discussions on what life skills are and how an understanding of them can help enrich people's lives.

References

American College Testing (1993). *Performing a National Job Analysis Study: Technical Report*. Iowa City, IA: Author.

Australian Education Council and Ministers for Vocational Education, Employment, and Training (1993). *Putting General Education to Work. The Key Competencies Report*. Author.

Berry, J. (1974). “Radical cultural relativism and the concept of intelligence.” In J.W. Berry and P.R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural psychology* (pp. 225-229). London: Methuen.

Carnevale, A.P., Gainer, L.J., and Meltzer A.S. (1988). *Workplace Basics: The Essential Skills Employers Want*. Washington, DC: American Society for Training and Development.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Feuerstein, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability.* Baltimore, MD: University Park.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic.

Greenfield, P.M. (1997). “You can’t take it with you: Why ability assessments don’t cross cultures.” *American Psychologist, 52,* 1115-1124.

Guilford, J.P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Gustafsson, J.E. (1984). “A unifying model for the structure of intellectual abilities.” *Intelligence, 8,* 179-203.

Horn, J.L. (1994). “Fluid and crystallized intelligence, theory of.” In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 443-451). New York: Macmillan.

Hunt, E. (1980). “Intelligence as an information-processing concept.” *British Journal of Psychology, 71,* 449-474.

Jensen, A.R. (1998). *The* g *factor*. Greenwich, CT: Greenwood.

Jones, S. (1996). *Background for Canadian Basic Job Skills Test*, Human Resources Development Canada and Statistics Canada.

Laboratory of Comparative Human Cognition (1982). “Culture and intelligence.” In R.J. Sternberg (Ed.), *Handbook of intelligence* (pp. 642-719). New York: Cambridge University Press.

Matarazzo, J. (1994). “Biological measures of intelligence.” In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 193-200). New York: Macmillan.

Mikulekcy, L. (1982) “Job Literacy: The Relationship Between School Preparation and Workplace Actuality.” *Reading Research Quarterly* 17:400419.

Pestillo, P., and Yokich, S. (1988) *A Michigan Employability Profile*. Detroit: Governor’s Commission on Jobs and Economic Development.

Piaget, J. (1972). *The psychology of intelligence.* Totowa, NJ: Littlefield Adams.

Premier’s Council (1990). *People and Skills in the New Global Economy.* Toronto: Premier’s Council.

Saterfiel, T.H. (1993). “Work Keys: Strengthening the School-to-Work Transition.” Paper presented to the National Council on Measurement in Education, Atlanta.

Scribner, S. (1984). “Studying working intelligence.” In B. Rogoff and J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9-40). Cambridge, MA: Harvard University Press.

Secretary’s Commission on Achieving Necessary Skills (1991). *What Work Requires of Schools. A SCANS Report for America 2000: Executive Summary*. Washington, DC: The Secretary’s Commission on Achieving Necessary Skills, US Department of Labor.

Spearman, C. (1904). “General intelligence, objectively determined and measured.” *American Journal of Psychology, 15,* 201-293.

Spearman, C. (1923). *The nature of ‘intelligence’ and the principles of cognition* (2nd ed.). London: Macmillan. (1923 edition reprinted in 1973 by Arno Press, New York).

Sternberg, R.J. (1983). “Components of human intelligence.” *Cognition, 15*, 1-48.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York: Cambridge University Press.

Sternberg, R.J. (1997a). *Successful intelligence*. New York: Plume.

Sternberg, R.J. (1997b). “Successful intelligence: a broader view of who is smart in school and in life.” *International Schools Journal, 17,* 19-31.

Sternberg, R.J., and Detterman, D.K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.

Sternberg, R.J., and Kaufman, J.C. (1998). “Human abilities.” *Annual Review of Psychology, 49,* 479-502.

Sternberg, R.J., and Lubart, T.I. (1995). “Defying the crowd: Cultivating creativity in a culture of conformity.” New York: Free Press.

Sternberg, R.J., Wagner, R.K., Williams, W.M., and Horvath, J.A. (1995). “Testing common sense.” *American Psychologist, 50,* 912-927.

Sticht, T.G. (1978). *Literacy and Vocational Competence. Occasional Paper No. 39.* Columbus, OH: The National Center for Research in Vocational Education, Ohio State University. ERIC Document Reproduction Service No. ED 181 329.

Taylor, M.C., and Lewe, G.R. (1990). Basic Skills Training: A Launchpad for Success in the Workplace. Literacy Task Analysis Project, Interim Report. Ottawa: Adult Basic Education Department, Algonquin College.

Thurstone, L.L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Thurstone, L.L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Appendix A

Workplace Skill Lists

1. Skills triangle from people and skills in the new global economy

- Not Portable Workplace Skills:
  - Firm and Job Specific Skills
- Portable Workplace Skills:
  - Generic Technical
  - Analytic Problem Solving
  - Workplace Interpersonal
- Portable Basic Skills:
  - Motor Skills
  - Mathematics
  - Reading and Writing
  - Ability to Learn
  - Communications

2. Key competencies from putting general education to work: the key competencies report

Collecting, Analyzing, and Organizing Information	The capacity to locate, sift and sort information in order to select what is required and present it in a useful way, and evaluate both the information itself and the sources and methods used to obtain it.
Communicating Ideas and Information	The capacity to communicate effectively with others using the range of spoken, written, graphic, and other non-verbal means of expression.
Planning and Organizing Activities	The capacity to plan and organize one's own work activities, including making good use of time and resources, sorting out priorities and monitoring one's own performance.
Working with Others and in Teams	The capacity to interact effectively with other people both on a one-to-one basis and in groups, including understanding and responding to the needs of a client and working effectively as a member of a team to achieve a shared goal.
Using Mathematical Ideas and Techniques	The capacity to use mathematical ideas, such as number and space, and techniques, such as estimation and approximation, for practical purposes.
Solving Problems	The capacity to apply problem-solving strategies in purposeful ways, both in situations where the problem and the desired solution are clearly evident and in situations requiring critical thinking and creative approaches to achieve outcomes.
Using Technology	The capacity to apply technology, combining the physical and sensory skills needed to operate equipment with the understanding of scientific and technological principles needed to explore and adapt systems.

3. Michigan employability skills profile

Academic skills

- Read and understand written materials
- Understand charts and graphs
- Understand basic math
- Use mathematics to solve problems
- Use research and library skills
- Use specialized knowledge and skills to get a job done
- Use tools and equipment
- Speak in the language in which business is conducted
- Write in the language in which business is conducted
- Use scientific method to solve problems

Personal management skills

- Attend school/work daily and on time
- Meet school/work deadlines
- Develop career plans
- Know personal strengths and weaknesses
- Demonstrate self-control
- Pay attention to details
- Follow written and oral instructions
- Follow written and oral directions
- Work without supervision
- Learn new skills
- Identify and suggest new ways to get the job done

Teamwork skills

- Actively participate in a group
- Know the group's rules and values
- Listen to other group members
- Express ideas to other group members
- Be sensitive to the group members' ideas and views
- Be willing to compromise if necessary to best accomplish the goal
- Be a leader to compromise if necessary to best accomplish the goal
- Work in changing settings and with people of differing backgrounds

4. Conference board employability skills profile

- Academic Skills
  - Communicate
  - Think
  - Learn
- Personal Management Skills
  - Positive Attitudes and Behaviours
  - Responsibility
  - Adaptability
- Teamwork Skills
  - Work with Others

5. Skills list from workplace basics: The skills employers want

- 3R's, Reading, Writing, Computation
- Learning to learn
- Communication: Listening and Oral Communication
- Creative Thinking/Problem Solving
- Interpersonal/Negotiation/Teamwork
- Self-Esteem/Goal Setting-Motivation/Personal and Career Development
- Organizational Effectiveness/Leadership

6. American college testing *workKeys* specifications

- Locating and Using Information
- Reading for Information
- Listening
- Writing
- Applied Mathematics
- Teamwork
- Applied Technology (assessed as problem-solving skills)
- Observation

7. SCANS competencies and foundation skills

Workplace competencies effective workers can productively use:

Resources	They know how to allocate time, money, materials, space, and staff.
Interpersonal Skills	They can work on teams, teach others, serve customers, lead, negotiate, and work well with people from culturally diverse backgrounds.
Information	They can acquire and evaluate data, organize and maintain files, interpret and communicate, and use computers to process information.
Systems	They understand social, organizational, and technological systems; they can monitor and correct performance; and they can design or improve systems.
Technology	They can select equipment and tools, apply technology to specific tasks, and maintain and troubleshoot equipment.

Foundations skills competent workers in the high-performance workplace need:

Basic Skills	Reading, writing, arithmetic and mathematics, speaking, and listening.
Thinking Skills	The ability to learn, to reason, to think creatively, to make decisions, and to solve problems.
Personal Qualities	Individual responsibility, self-esteem and self-management, sociability, and integrity.

8. NCVQ core skills

- Communicate
    - Discussions
    - Write
    - Use images
    - Read and respond
- Application of Number
- Problem Solving
- Information Technology
- Personal Skills
    - Improving own learning and performance
    - Working with others

9. Essential skills research project

- Reading Skills
    - Textual Materials
    - Graphical Materials
- Writing Skills
- Numeracy Skills
- Psychomotor Skills
- Oral Communication Skills
- Thinking Skills
    - Problem Solving
    - Decision making
    - Planning/Organizing Job Tasks
- Working with Others
- Computer Skills
- Continuous Learning

Part II

Theoretical Frameworks for Specific Domains Included in ALL

This Part includes three chapters that provide assessment frameworks for the four skill domains that met the criteria set out for inclusion in the international comparative assessment. These chapters also document the processes that were used to develop and validate the assessment instruments and presents data related to validity and reliability. Chapter 4 presents the framework for prose, document and quantitative literacy. This framework covers the three domains that were included in the International Adult Literacy Survey (IALS). Quantitative literacy was replaced by Numeracy for the ALL study. This framework served as a model for other frameworks. Chapter 5 presents the framework for numeracy, and, chapter 6 presents the framework for problem solving.

Chapter 4

Prose Literacy, Document Literacy and Quantitive Literacy: Understanding What Was Measured in IALS and ALL

Irwin Kirsch *Educational Testing Service Princeton, NJ*

Materials adapted from "*The International Adult Literacy Survey (IALS): Understanding What Was Measured.* Research Report, Kirsch, Irwin, Educational Testing Service, December 2001." Used by permission of Educational Testing Service, the copyright owner.

Table of Contents

Abstract

This paper offers a framework that has been used for both developing the tasks used to measure literacy and for understanding the meaning of what has been reported with respect to the comparative literacy proficiencies of adults in participating countries. The framework consists of six parts that represent a logical sequence of steps, from needing to define and represent a particular domain of interest, to identifying and operationalizing characteristics used to construct items, to providing an empirical basis for interpreting results. The various parts of the framework are seen as important in that they help to provide a deeper understanding of the construct of literacy and the various processes associated with it. A processing model is proposed and variables associated with performance on the literacy tasks are identified and verified through regression analyses. These variables are shown to account for between 79% and 89% of the variance in task difficulty. Collectively, these process variables provide a means for moving away from interpreting performance on large-scale surveys in terms of discrete tasks or a single number, toward identifying levels of performance that have generalizability across pools of tasks and toward what Messick (1989) has called a higher level of measurement.

Key words: assessment design, construct validity, evidence centered assessment design, framework, large-scale assessment, literacy/reading, test interpretation

Introduction

The International Adult Literacy Survey (IALS) was the first-ever comparative survey of adults designed to profile and explore the literacy distributions among participating countries. It was a collaborative effort involving several international organizations, intergovernmental agencies, and national governments. In 2000, a final report was released (Organization for Economic Co-operation and Development and Statistics Canada, 2000), which stated that "by 1998, the survey had covered 10.3 percent of the world population and 51.6 percent of the world GDP" (p.87).

Who are the constituencies that are likely to use the data from the IALS once they have been collected and analyzed? It is expected that many individuals, including researchers, practitioners, and individual citizens within each of the participating countries, will read the survey results and make use of the data for a variety of purposes. Yet, the primary reason for developing and conducting this large-scale international assessment is to provide empirically grounded interpretations upon which to inform policy decisions. This places the IALS in the context of policy research. In their classic volume on this topic, Lerner and Lasswell (1951) argued that the appropriate role for policy research is not to define policy; rather, it is to establish a body of evidence from which informed judgments can be made. Messick (1987) extended this thinking to the area of large-scale assessments and noted that, in order to appropriately fulfill this function, assessments should exhibit three key features: *relevance*, *comparability*, and *interpretability*.

Relevance refers to the capability for measuring diverse background and program information to illuminate context effects and treatment or process differences. Both IALS and ALL developed and administered an extensive questionnaire covering a wide range of issues that will be used to identify characteristics that are correlated with performance and that may differ across a variety of language and cultural backgrounds.

Comparability deals with the capacity to provide data or measures that are commensurable across time periods and across populations of interest. Complex sampling, scaling, and translation procedures are being implemented to help ensure that common metrics will exist across participating countries so that appropriate comparisons can be made between countries and among major subpopulations of interest within a country. These comparisons are important both in this initial survey and in future assessments where new countries may join the survey and want to be placed onto existing scales, or where participating countries may want to measure trends in the distributions of skills among various subpopulations of interest.

Interpretability focuses on collecting evidence that will enhance the understanding and interpretation of what is being measured. In some assessments, the meaning of what is being measured is constructed by examining performance on individual tasks, or by assuming it is inherent in the label that is used to organize one or more sets of tasks—for example, reading comprehension or critical thinking. All too often assessments focus on rank ordering populations or countries by comparing mean scores or distributions. These data tell us that people differ without telling us how they differ. One of the stated goals in the IALS and ALL studies is to try to address the issue of interpretability not only by reporting that countries, groups, or individuals differ in their proficiencies, but also by developing an interpretative scheme for reporting how they differ.

In considering the development of the literacy framework, a set of necessary components has been identified:

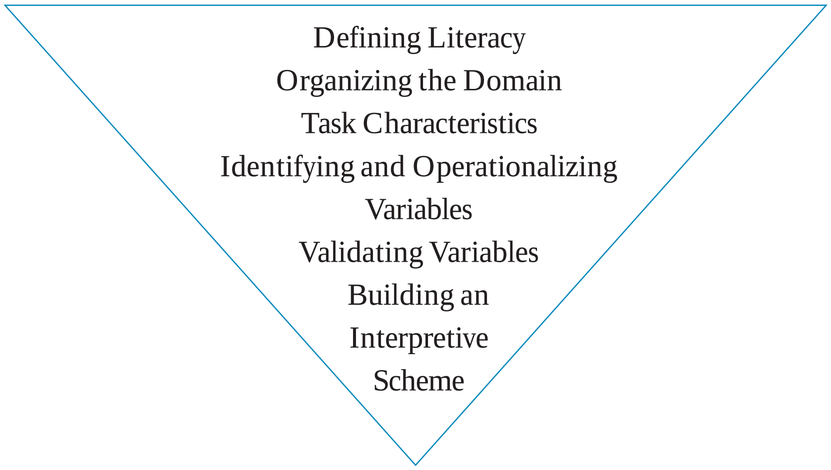
- A framework should begin with a general definition or statement of purpose—one that guides the rationale for the survey and what should be measured.
- A framework should identify various task characteristics and indicate how these characteristics will be used in constructing the tasks.
- Variables associated with each task characteristic should be specified, and research should be conducted to show which of these variables account for large percentages of the variance in the distribution of tasks along a continuum or scale. Variables that appear to have the largest impact on this variance should be used to create an interpretative scheme. This is a crucial step in the process of measurement and validation.

While the chief benefit of constructing and validating a framework for literacy is improved measurement, a number of other potential benefits are also evident. Namely:

- A framework provides a common language and a vehicle for discussing the definition of the skill area.
- Such a discussion allows us to build consensus around the framework and measurement goals.
- An analysis of the kinds of knowledge and skills associated with successful performance provides a basis for establishing standards or levels of proficiency. As we increase our understanding of what is being measured and our ability to interpret scores along a particular scale, we have an empirical basis for communicating a richer body of information to various constituencies.
- Identifying and understanding particular variables that underlie successful performance further our ability to evaluate what is being measured and to make changes to the measurement over time.
- Linking research, assessment, and public policy promotes not only the continued development and use of the survey, but also understanding of what it is measuring.

Overview of the framework

While there are many approaches one could take to develop a framework for measuring a particular skill area, the diagram shown here represents a process that has been used to construct and interpret the literacy tasks for the National Adult Literacy Survey (NALS) (Kirsch, Jungeblut, Jenkins, and Kolstad, 1993) and for the IALS (OECD and Human Resources Development Canada [HRDC], 1997; OECD and Statistics Canada, 1995; OECD and Statistics Canada, 2000). This process is also being used to develop the reading literacy measure for the Programme for International Student Assessment (PISA) (OECD, 1999). The diagram shown here represents a process that consists of six parts. These six parts represent a logical sequence of steps that should be addressed, from needing to define a particular skill area, to having specifications for constructing items, to providing an empirically based interpretation of the scores that are obtained.



Part 1 of the framework focuses on the working definition for literacy, along with some of the assumptions that underlie it. In doing so, the definition sets the boundaries for what the survey seeks to measure as well as what it will not measure. Part 2 provides a discussion on how we may choose to organize the set of tasks that are constructed to report to policymakers and researchers on the distribution of a particular skill in the population. Determining how to report the data should incorporate statistical, conceptual, and political considerations. Part 3 deals with the identification of a set of key characteristics that will be manipulated by developers when constructing tasks for a particular skill area. Part 4 identifies and begins to define the variables associated with the set of key characteristics that will be used in test construction. These definitions are based on the existing literature and on experience with building and conducting other large-scale assessments. Part 5 lays out a procedure for validating the variables and for assessing the contribution each makes toward understanding task difficulty across the various participating countries. The final part, Part 6, discusses how an interpretative scheme was built using the variables that have been shown through the research in Part 5 to account for task difficulty and student performance.

Defining literacy

Definitions of reading and literacy have changed over time in parallel with changes in our society, economy, and culture. The growing acceptance of the importance of lifelong learning has expanded the views and demands of reading and literacy. Literacy is no longer seen as an ability that is developed during the early school years, but is instead viewed as an advancing set of skills, knowledge, and strategies that individuals build on throughout their lives in various contexts and through interaction with their peers and with the larger communities in which they participate.

Historians remind us that the types and levels of literacy skills necessary for economic participation, citizenship, parenting, and individual advancement in 1800 were different from those required by 1900 and from those required in the year 2000 and beyond. We live in a technologically advancing world, where both the number and types of written materials are growing and where increasing numbers of citizens are expected to use information from these materials in new and more complex ways.

As Resnick and Resnick (1977) point out, literacy in its earliest form consisted of little more than signing one's name. It was not until much later that fluent oral reading became important, and not until the 20th century that reading to gain information was given primary emphasis. Standardized tests became fashionable and reading-grade-level scores became the focus of attention. Through the use of these instruments the term *literacy* has implied the acquisition of intellectual skills associated with basic academic competencies associated with reading and writing. Standards for literacy increased over the decades, from being able to read at a fourth-grade level, to reading at an eighth-grade level, and then by the early '70s, to a 12th-grade level. These measures came under increasing criticism, however, because they did not provide specific information about the kinds of competencies that given levels of literacy imply. Perhaps more important was the recognition that literacy relates not to some arbitrary standard for the purpose of categorizing people as literate or illiterate, but to what people can do with printed and written materials and how these skills relate to a host of social needs. As Beach and Appleman (1984) noted,

> The often heard charge, Johnny can't read is a little like saying Johnny can't cook. Johnny may be able to read the directions for constructing a radio kit, but not a Henry James novel, just as Johnny may be able to fry an egg but not cook Peking duck. In discussing reading in the schools, we must recognize that reading involves as wide a range of different types of texts as there are types of food. And, to imply, as does the slogan, "Johnny can't read," that reading is a single skill suited to all types of texts does not do justice to the range of reading types.

Thus, the multifaceted nature of literacy had often been glossed over through the use of grade-level equivalent scores.

It was from this multifaceted perspective that several large-scale assessments of literacy were conducted in Australia (Wickert, 1989), Canada (Montigny, Kelly, and Jones, 1991), and the United States (Kirsch and Jungeblut, 1986; Kirsch et al., 1993).

In 1992, the Organization for Economic Co-operation and Development (OECD) (OECD and Statistics Canada, 1992) concluded that low literacy levels were a serious threat to economic performance and social cohesion on an international level. But a broader understanding of literacy problems across industrialized nations—and consequent lessons for policymakers—was hindered due to a lack of comparable international data. Statistics Canada and Educational Testing Service (ETS) teamed up to build and deliver an international comparative study. After some discussion and debate, the framework and methodology used in NALS was applied to the first large-scale International Adult Literacy Survey.

NALS, which was funded by the National Center for Education Statistics (NCES) as part of its overall assessment program in adult literacy, was the largest and most comprehensive study of adult literacy ever conducted in the United States. Like all large-scale assessments funded by the NCES, NALS was guided by a committee, which was comprised of a group of nationally recognized scholars, practitioners, and administrators who adopted the following definition of literacy:

Literacy is using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.

This definition captures the initial work of the committee guiding the development of the assessment, and provides the basis for creating other aspects of the framework to be discussed. It also carries several assumptions made by panel members; thus, it is important to consider various parts of this definition in turn.

Literacy is …

The term *literacy* is used in preference to *reading* because it is likely to convey more precisely to a nonexpert audience what the survey is measuring. *Reading* is often understood as simply decoding or reading aloud, whereas the intention of the adult surveys is to measure something broader and deeper. Researchers studying literacy within particular contexts noted that different cultures and groups may value different kinds of literacy practices (Heath, 1980; Sticht, 1975; Szwed, 1981). Heath, for example, found that uses for reading could be described in terms of instrumental, social interactional, news-related, memory supportive, substitutes for oral messages, provision of a permanent record, and personal confirmation. The fact that people read different materials for different purposes implies a range of proficiencies that may not be well captured by signing one's name, completing a certain number of years of schooling, or scoring at an eighth-grade level on a test of academic reading comprehension.

… using printed and written information

This phrase draws attention to the fact that panel members view literacy not as a set of isolated skills associated with reading and writing, but more importantly as the application of those skills for specific purposes in specific contexts. When literacy is studied within varying contexts, diversity becomes its hallmark. First, people engage in literacy behaviors for a variety of uses or purposes (Cook-Gumperz and Gumperz, 1981; Heath, 1980; Mikulecky, 1982; Sticht, 1978). These uses vary across contexts (Heath, 1980; Venezky, 1983) and among people within the same context (Kirsch and Guthrie, 1984a). This variation in use leads to an interaction with a broad range of materials that have qualitatively different linguistic forms (Diehl, 1980; Jacob, 1982; Miller, 1982). In some cases, these different types of literacy tasks have been associated with different cognitive strategies or reading behaviors (Crandall, 1981; Kirsch and Guthrie, 1984b; Scribner and Cole, 1981; Sticht, 1978, 1982).

… to function in society, to achieve one's goals, and to develop one's knowledge and potential.

This phrase is meant to capture the full scope of situations in which literacy plays a role in the lives of adults, from private to public, from school to work, to lifelong learning and active citizenship. "[T]o achieve one's goals and to develop one's knowledge and potential" points to the view that literacy enables the fulfillment of individual aspirations—both defined ones such as graduation or obtaining a job, and those less defined and less immediate, which extend and enrich one's personal life. The phrase "to function in society" is meant to acknowledge that literacy provides individuals with a means of contributing to, as well as benefiting from, society. Literacy skills are generally recognized as important for nations to maintain or improve their standard of living and to compete in an increasingly global marketplace. Yet, they are equally as important for individual participation in technologically advancing societies with their formal institutions, complex legal systems, and large government programs.

Organizing the domain

Having defined the domain of literacy and having laid out the set of assumptions that were made in developing the definition, it is important to think about how to organize the domain. This organization needs to focus on how to report the scores that result from administering a pool of literacy tasks. This is an important issue because how the domain is organized can affect test design. Because some believe that reading is not a single, one-dimensional skill, literacy is not necessarily best represented by a single scale or single score along that scale. Yet determining how many and which scales should be used for reporting literacy scores is crucial for ensuring that sufficient numbers of tasks are developed to define and interpret these scales adequately.

Different perspectives can be used to help organize a domain of tasks. Traditionally, literacy skills have been categorized by modality into reading, writing, speaking, and listening. Reading and writing are sometimes combined, as they are thought to require similar processes, and speaking and listening are often grouped in terms of being too costly and difficult to assess. Thus, they were not included in the survey. Committee members also wanted to include basic arithmetic calculations as part of the assessment since adults are often required to use printed information that involves these skills. As a result, this aspect of literacy was also included in the surveys.

Work in the area of context of literacy clearly provides one possible organizing principle for what may appear to be a disparate set of literacy tasks. There is the familiar academic or school context (dealing primarily with prose or connected discourse) contrasted with nonschool or "everyday life" contexts. And the nonschool contexts can be subdivided into the work-related and home-related tasks. However, it is operationally difficult to separate tasks along these latter dimensions since the work and home categories are not mutually exclusive in terms of the literacy tasks engaged in.

Another organizing principle of some appeal involves categorizing literacy tasks in terms of the types of materials or formats in which they occur, and to examine the associated purposes or uses both within and across materials. The appeal for this type of organizational scheme stems from research literature suggesting that different materials or formats are associated with different contexts and that a significant proportion of adult reading tasks in the context of work involve documents (Jacob, 1982; Kirsch and Guthrie, 1984a; Sticht, 1975)—graphs, charts, forms, and the like—rather than prose. Frequently, these documents are embedded in the contexts of home or work and community, as contrasted with prose, which is most frequently associated with school or academia. Moreover, different materials and formats are often associated with different purposes, and these purposes are frequently associated with different reading strategies. This line of reasoning led to distinctions such as Sticht's "reading to do" and "reading to learn."

As another instance reflecting similar distinctions, the National Assessment of Educational Progress (NAEP) (1972) came to aggregate reading exercises in terms of "themes"—word meanings, visual aids, written directions, reference materials, significant facts, main ideas, inferences, and critical reading. The areas of reference materials and significant facts were among those in which young adults aged 26-35 performed better than did in-school 17-year-olds, while in-school 17-year-olds performed higher than young adults in inferences and critical reading. These and other NAEP results suggest the utility of a priori classifications that allow for the examination of differential performance for subgroups both within a single assessment and across groups over time.

In the end, a compromise was reached among the various organizing concepts that was felt to reflect a number of salient notions from the literature. Three scales were hypothesized—a prose literacy scale, a document literacy scale, and a quantitative literacy scale. In this way, it is possible to acknowledge that the structure of prose passages are qualitatively different from the structures associated with documents such as charts, tables, schedules, and the like, and to provide for a separate scale for those tasks involving the processing of printed information in combination with arithmetic operations.

The original data from the NAEP Young Adult Literacy Survey (YALS) was subjected to factor analysis to explore dimensionality (Kirsch and Jungeblut, 1986). Following the logic of Cattell's scree test (1966), the breaks in the pattern of latent roots indicated at least three salient factors, with the possibility of as many as five additional factors. Analysis of parallel random data reinforced the judgment that a three-factor solution was appropriate. However, for exploratory purposes, three separate analyses were conducted: In one analysis eight factors were retained and rotated for interpretation; in another, five factors were retained; and, in the final analysis, three factors were retained for rotation and interpretation.

In each instance, the factors were rotated to orthogonal simple structure by the varimax procedure and to oblique simple structure by the DAPPER method (Tucker and Finkbeiner, 1981). Tasks loading highest on the first and largest factor seemed to rely heavily on prose comprehension, tasks loading highest on the second factor seemed to reflect skills in using documents, while tasks loading highest on the third factor required the application of arithmetic operations.

Interpretation of the five- and eight-factor solutions was much less clear. Although each revealed three major factors reflecting prose, document, and quantitative operations, for the most part these rotated solutions provide interesting clues for possible task modification and for future item development, rather than clear-cut implications for scaling the existing data. That is, if desired, one could devise a new set of tasks that could isolate a factor reflecting the importance of procedural knowledge as it might apply, for example, to entering and using information in forms. Alternatively, one might prefer to restrict the impact of this type of knowledge by eliminating this type of task from the assessment. Thus, the empirical data provided by the YALS tended not only to support the a priori judgment for the three literacy scales but also suggested ways in which the assessment could be broadened. It is important to keep in mind that the three literacy scales are not the only salient dimensions of literacy per se. These dimensions are likely to shift as a function of different definitions and different perspectives on literacy.

More recent advisory committees involved with NALS and IALS have agreed that literacy should not be measured along a single continuum and have chosen to adopt the general definition and three scales defined here. These committees further recommended that new literacy tasks, which were constructed for each of these assessments, should be developed to enhance the three existing scales, and that these new tasks should continue to use open-ended simulation tasks rather than multiple-choice questions and to emphasize measuring a broad range of information-processing skills covering a variety of contexts.

Identifying task characteristics

Almond and Mislevy (1998) note that variables can take on one of five roles in an assessment or test. They can be used to limit the scope of the assessment, characterize features that should be used for constructing tasks, control the assembly of tasks into booklets or test forms, characterize examinees' performance on or responses to tasks, or help to characterize aspects of competencies or proficiencies. Some of these variables can be used both to help in the construction of tasks and the understanding of competencies, as well as in the characterization of performance. A finite number of characteristics are likely to influence students' performance on a set of literacy tasks, and these can be taken into account when constructing or scoring the tasks. These characteristics, which are thought to be important components of the literacy process, were manipulated in the development of tasks for IALS. These characteristics include:

- *Adult Contexts/Content*. Since adults do not read written or printed materials in a vacuum, but read within a particular context or for a particular purpose, materials for the literacy assessment are selected that represent a variety of contexts and contents. This helps ensure that no single group of adults is either advantaged or disadvantaged due to the context or content included in the assessment.

- *Materials/Texts*. While no one would doubt that a literacy assessment should include a range of material, what is critical to the design and interpretation of the scores that are produced are the range and specific features of the text material that are included in constructing the tasks. Thus, a broad range of both prose and document text types are included in this survey.
- *Processes/Strategies*. This refers to the characteristics of the questions and directives that are given to adults for their response. Generally speaking, the questions and directives will refer to a goal or purpose the readers are asked to assume while they are reading and interacting with texts, and relate to one or more strategies that the reader is likely to use in producing their response.

Identifying and operationalizing the variables

In order to use these three main task characteristics in designing the assessment and, later, in interpreting the results, the task characteristics need to be operationalized. That is, various values that each of these characteristics can take on must be specified. This will allow item developers to categorize the materials they are working with and the questions and directives they construct so that they can be used in the reporting of the results. These variables can also be used to specify what proportions of the assessment ought to come from each category.

Context/Content

Materials that are selected for inclusion in the assessment need to represent a broad range of contexts and contents so that no single group is advantaged or disadvantaged in terms of familiarity or exposure. Six adult context/content categories have been identified as follows:

- *Home and family* includes materials dealing with interpersonal relationships, personal finance, housing, and insurance.
- *Health and safety* includes materials dealing with drugs and alcohol, disease prevention and treatment, safety and accident prevention, first aid, emergencies, and staying healthy.
- *Community and citizenship* includes materials dealing with community resources and staying informed.
- *Consumer economics* includes materials dealing with credit and banking, savings, advertising, making purchases, and maintaining personal possessions.
- *Work* includes materials that deal in general with various occupations but not job-specific texts, finding employment, finance, and being on the job.
- *Leisure and recreation* includes materials involving travel, recreational activities, and restaurants.

It is important to note that with respect to this variable, an attempt should be made to include as broad a range as possible across the six contexts, as well as to select universally relevant materials. Following this procedure will help to ensure that the content and materials that are included in the assessment are not so specialized as to be familiar only to certain groups and that any disadvantages for people with limited background knowledge might be minimized.

Materials/Texts

Reading requires something for the reader to read. In an assessment, that something—a text—must be coherent within itself. That is, the text must be able to stand alone without requiring additional printed material. While it is obvious that there are many different kinds of texts and that any assessment should include a broad range of them, it is not so obvious that there is an ideal categorization of text types. There are any number of proposals as to the appropriate categories, many of them created for practical rather than theoretical purposes. All of them share the fact that no particular physical text seems to fit easily into only one category. For example, a chapter in a textbook might include some definitions (often identified as a text type), some instructions on how to solve particular problems (yet another text type), a brief historical narrative of the discovery of the solution (still another text type), and descriptions of some typical objects involved in the solution (a fourth text type).

It might be thought that a definition, for example, could be extracted and treated as a single text for assessment purposes. But this would remove the definition from the context, create an artificial text type (definitions almost never occur alone, except in dictionaries), and not allow item writers to create tasks that deal with reading activities that require integrating information from a definition with information from instructions.

A more important classification of texts, and one at the heart of this assessment, is the distinction between continuous and noncontinuous texts. Continuous texts are typically composed of sentences that are, in turn, organized into paragraphs. These may be fit into even larger structures such as sections, chapters, and books. Noncontinuous texts are most frequently organized in matrix format, based on combinations of lists.

Continuous texts

Conventionally, continuous texts are formed of sentences organized into paragraphs. In these texts, organization occurs by paragraph setting, indentation, and the breakdown of text into a hierarchy signaled by headings that help the reader recognize the organization of the text. Text types are standard ways of organizing the contents of and author's purpose for continuous texts.[1]

1. *Description* is the type of text where the information refers to properties of objects *in space*. Descriptive texts typically provide an answer to *what* questions.
2. *Narration* is the type of text where the information refers to properties of objects *in time*. Narration texts typically provide answers to *when*, or *in what sequence,* questions.
3. *Exposition* is the type of text in which the information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analyzed. The text provides an explanation of how the component elements interrelate in a meaningful whole and often answers *how* questions.
4. *Argumentation* is the type of text that presents propositions as to the relationship among concepts or other propositions. Argument texts often answer *why* questions. Another important subclassification of argument texts are persuasive texts.
5. *Instruction* (sometimes called *injunction*) is the type of text that provides directions on what to do.
6. *Document* or *record* is a text that is designed to standardize and conserve information. It can be characterized by highly formalized textual and formatting features.

7. *Hypertext* is a set of text slots linked together in such a way that the units can be read in different sequences, allowing readers to follow various routes to the information.

Noncontinuous texts

Noncontinuous texts are organized differently than continuous texts and so allow the reader to employ different strategies for entering and extracting information from them. On the surface, these texts appear to have many different organizational patterns or formats, ranging from tables and schedules to charts and graphs, and from maps to forms. However, the organizational pattern for these types of texts, which Mosenthal and Kirsch (1998) refer to as documents, is said to have one of four basic structures: a simple list, a combined list, an intersected list, or a nested list. Together, these four types of documents make up what they have called matrix documents, or noncontinuous texts with clearly defined rows and columns. They are also closely related to other noncontinuous texts that these authors refer to as graphic, locative, and entry documents.[2]

1. *Matrix Documents*. This set of noncontinuous text consists of four types of increasingly complex documents that have simple lists as their basic unit. A simple list consists of a label and two or more items, where the label serves as the organizing category and the items all share at least one feature with the other items in the list. Next are combined lists, which consist of two or more simple lists. One list in a combined list is always primary and, as such, is ordered to facilitate looking up information within the list and locating parallel information within the other lists. Intersected lists are the third type of matrix document and comprise exactly three lists. Two of the lists form a row and column defining the cells of the third or intersected list. The fourth and most complex type of matrix document is the nested list. In order to economize on space, as well as to display comparative information, designers sometimes combine two or more intersecting lists to form a nested list. In a nested list, one type of information will be repeated in each of the intersecting lists. The intersecting list of unemployment rates, for example, may have separate entries under each month for males and females; in this case, gender would be nested under month.
2. *Graphic Documents*. A major function of graphic documents is to provide a succinct visual summary of quantitative information. Included in this group of documents or noncontinuous texts are pie charts, bar charts, and line graphs. While these appear to be very different types of documents on the surface, they all derive or can be transformed into either a combined, intersecting, or nested list.
3. *Locative Documents*. Like graphic documents, locative documents or maps portray information visually. Unlike graphic documents that display quantitative information, maps either portray the location of persons, places, or things in space, or depict characteristics of different geographic regions (e.g., types of vegetation or characteristics of a population).
4. *Entry Documents*. In matrix and graphic documents, the author provides the information that must be read and used. In contrast, entry documents or forms require the reader to provide information that can range from very simple to complex. For example, the reader may be asked to simply check a box; write a single word, number, or phrase; or construct a series of phrases or sentences. Generally speaking, forms provide the reader with a label or category for which the reader is asked to provide specifics.

5. *Combination Documents*. It is important to keep in mind that some displays, especially graphic documents, rely on the use of other documents for their interpretation. Maps and graphs, for instance, often include legends that display important information that must be read and understood. In addition, designers sometimes include more than one document for display or comparative purposes.

Processes/Strategies

This task characteristic refers to the way in which examinees process text to respond correctly to a question or directive. It includes the processes used to relate information in the question (the given information) to the necessary information in the text (the new information), as well as the processes needed to either identify or construct the correct response from the information available. Three variables in the reading/literacy research used to investigate tasks from national and international surveys will be considered here. These are: type of match, type of information requested, and plausibility of distracting information. They are briefly described here. They are characterized through a discussion of exemplary tasks in the next section and fully operationalized in the appendix at the end of this paper.

Type of match

Four types of matching strategies were identified: locating, cycling, integrating, and generating. Locating tasks require examinees to match one or more features of information stated in the question to either identical or synonymous information provided in the text. Cycling tasks also require examinees to match one or more features of information, but unlike locating tasks, they require respondents to engage in a series of feature matches to satisfy conditions stated in the question. Integrating tasks require examinees to pull together two or more pieces of information from the text according to some type of specified relationship. For example, this relationship might call for examinees to identify similarities (i.e., make a comparison), differences (i.e., contrast), degree (i.e., smaller or larger), or cause-and-effect relationships. This information may be located within a single paragraph or it may appear in different paragraphs or sections of the text. In integrating information, examinees draw upon information categories provided in a question to locate the corresponding information in the text. They then relate the text information associated with these different categories based upon the relationship term specified in the question. In some cases, however, examinees must generate these categories and/or relationships before integrating the information stated in the text.

In addition to requiring examinees to apply one of these four strategies, the type of match between a question and the text is influenced by several other processing conditions that contribute to a task's overall difficulty. The first of these is the number of phrases that must be used in the search. Task difficulty increases with the amount of information in the question for which the examinee must search in the text. For instance, questions that consist of only one independent clause tend to be easier, on average, than those that contain several independent or dependent clauses. Difficulty also increases with the number of responses that examinees are asked to provide. Questions that request a single answer are easier than those that require three or more answers. Further, questions that specify the number of responses tend to be easier than those that do not. For example, a question that states, "List the three reasons . . ." would be easier than one that said, "List the reasons . . ." Tasks are also influenced by the degree to which examinees have to make inferences to match the given information in a question to corresponding information in the text, and to identify the requested information. An additive scoring model defining type of match for prose and document literacy tasks is provided in Appendix A.

Type of information requested

This refers to the kinds of information that readers identify to answer a test question successfully. The more concrete the requested information, the easier the task is judged to be. In previous research based on large-scale assessments of adults' and children's literacy (Kirsch, Jungeblut, and Mosenthal, 1998; Kirsch and Mosenthal, 1994), the type of information variable was scored on a 5-point scale. A score of 1 represented information that was the most concrete and therefore the easiest to process, while a score of 5 represented information that was the most abstract and therefore the most difficult to process. For instance, questions that asked examinees to identify a person, animal, or thing (i.e., imaginable nouns) were said to request highly concrete information and were assigned a value of 1. Questions asking respondents to identify goals, conditions, or purposes were said to request more abstract types of information. Such tasks were judged to be more difficult and received a value of 3. Questions that required examinees to identify an "equivalent" were judged to be the most abstract and were assigned a value of 5. In such cases, the equivalent tended to be an unfamiliar term or phrase for which respondents had to infer a definition or interpretation from the text.

Plausibility of distractors

This concerns the extent to which information in the text shares one or more features with the information requested in the question but does not fully satisfy what has been requested. Tasks are judged to be easiest when no distractor information is present in the text. They tend to become more difficult as the number of distractors increases, as the distractors share more features with the correct response, and as the distractors appear in closer proximity to the correct response. For instance, tasks tend to be judged more difficult when one or more distractors meet some but not all of the conditions specified in the question and appear in a paragraph or section of text other than the one containing the correct answer. Tasks are judged to be most difficult when two or more distractors share most of the features with the correct response and appear in the same paragraph or node of information as the correct response.

At first glance, the skills involved in performing quantitative tasks might appear to be fundamentally different from those involved in processing prose and document tasks. An analysis of tasks along this scale shows, however, that processing printed information plays an important role in affecting the difficulty of quantitative tasks. In general, it appears that many individuals can perform single arithmetic operations using printed materials when both the numbers and operations are made explicit. Yet, when the numbers for these same operations must be extracted from materials that contain similar but irrelevant information, or when the operations must be inferred, the tasks become increasingly difficult.

As with the prose and document tasks, quantitative tasks require individuals to match information in a question or directive with information stated in one or more texts where a text could be either continuous or noncontinuous. In addition, quantitative tasks may require respondents to deal with plausible distractors when extracting information for an arithmetic operation. Individuals are also required to process some type of information. While type of information varies for the prose and document tasks, requested information is always an amount in quantitative tasks. Thus, the process variables for quantitative tasks are type of match and plausibility of distractors—like those defined for prose and document literacy—plus two additional variables that are unique to this scale. These are *type of calculation* and *operation specificity*. These two variables are briefly described here. They are more fully characterized through a discussion of exemplary tasks and fully operationalized in Appendix A.

Type of calculation

This variable includes both the type of arithmetic operation (addition, subtraction, multiplication, or division) required and whether that operation must be performed alone or in combination. Tasks involving multiplication and division tend to be more difficult than those requiring addition and subtraction, and tasks requiring two or more operations tend to be more difficult than tasks requiring only a singe operation. Codes for this variable ranged from 1 (easiest) to 5 (most difficult).

Operation specificity

This variable refers to the process of identifying and sometimes entering numbers into an arithmetic expression, including determining the appropriate operation to be performed. Tasks tend to be more difficult when the numbers that must be identified appear in a text and are neither in column format nor adjacent to each other. Tasks also tend to become more difficult when the operation(s) is not specified and when the wording in the question or directive does not contain an explicit semantic relationship statement such as “how many” or “calculate the difference.” The codes for operation specificity ranged from 1 (easiest) to 9 (most difficult) based on a set of additive rules reflecting the various facets described here and fully operationalized in Appendix A.

In previous surveys, the goal has been to develop pools of prose, document, and quantitative tasks that represent the range of contexts, texts, and processes outlined here, with no specific requirement for particular numbers of any type of task. The goal was to draw materials from a wide variety of adult contexts that represented a wide range of linguistic structures such as those outlined in this paper. With respect to continuous or prose texts, the focus has been on expository texts since much of what adults read for work and in their community is associated with this type of discourse. However, some surveys did include narratives and poetry in small numbers. In terms of processes/strategies, the goal was to engage adults in the full range of processes that might reasonably be associated with each type of material. That is, the goal was to use the framework to construct questions/directives that were thought to be authentic to the kinds of information someone might want to understand or use from a particular text.

Validating the variables

In a previous section, three task characteristics labeled *context*, *texts*, and *process/strategy* were introduced. It was followed by a section in which each task characteristic was operationalized into a number of variables. This part of the framework describes a procedure for validating the set of variables developed from these characteristics that have been shown to affect task performance and the placement of tasks along each of the reporting scales. This process borrows heavily from work that has been done in the area of adult literacy where several national and international surveys have reported data that followed this approach:

- The US Department of Labor’s Literacy Assessment (Kirsch and Jungeblut, 1992)

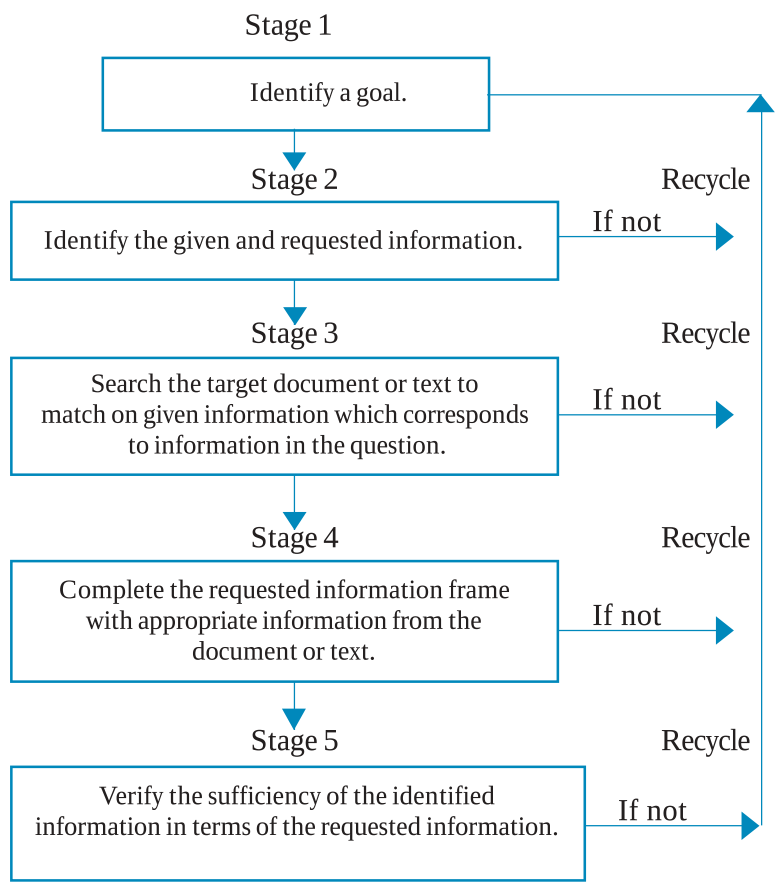
- The International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy Study (Kirsch and Mosenthal 1994)
- The National Adult Literacy Survey (Kirsch et al., 1993)

Reading tasks for these surveys were developed to represent a broad range of purposes for which students and adults read continuous and noncontinuous texts in both school and nonschool settings. To identify the variables contributing to task difficulty in each of the literacy domains, Kirsch and Mosenthal (Kirsch et al., 1998; Kirsch and Mosenthal, 1990) began by modeling the processes required to complete prose, document, and quantitative tasks in the literacy assessments. This model is shown in Figure 1 and grew out of earlier exploratory work (Fisher, 1981; Guthrie, 1988; Kirsch and Guthrie, 1984b).

In the first step, readers identify a goal or purpose for searching and processing a text or document. In a test or an instructional situation, questions and directives determine the primary purpose for interacting with a text or document, and therefore also determine the information that readers must process in order to complete a cognitive activity. In open-ended tasks, the reader's goal is to identify information in the text that meets the conditions set forth in the question or directive. In multiple-choice tasks, the reader's goal is to identify information in the text that meets the conditions set forth in the question or directive and then to select the best choice from a list of options (Kirsch and Mosenthal, 1994).

Figure 1

A model of prose and document processing in reading



In the second step, readers must distinguish between “given” and “requested” information in the question (Clark and Haviland, 1977; Mosenthal and Kirsch, 1991). Given information is presumed to be true, and it conditions the requested information. Requested information, on the other hand, is the specific information being sought.

In the third step, readers must search and read (or read and search) a text or document to identify the necessary information that corresponds with information provided in the question and, in the case of multiple-choice items, in the list of choices. In carrying out this search, several matches may be tried before one or more adequate matches are achieved. If a literal or synonymous match is made between requested or given information and corresponding text or document information, readers may proceed to the next step. If such a match is not deemed adequate, readers may choose to make a match based on a low- or high-level text-based inference or on prior knowledge; or readers may recycle to the first step.

This test-taking model of reading can be applied to multiple-choice as well as open-ended tasks. Based on this model, Kirsch and Mosenthal identified three variables as being among the best predictors of task difficulty for the prose and document scales. Two additional variables were constructed for the quantitative scales. These variables (type of requested information, type of match, plausibility of distractors, type of calculation, and operation specificity) were described in the previous section and are elaborated in Appendix A.

In order to understand how these variables interact with one another to affect the difficulty of items developed for the IALS, each literacy scale will be characterized in terms of several exemplary tasks. Next, these variables will be evaluated in terms of their contribution toward explaining the placement of literacy tasks along their respective scales.

Characterizing prose literacy tasks

There are 34 tasks ordered along the IALS 500-point prose literacy scale. These tasks range in difficulty value from 188 to 377. The easiest task (receiving a difficulty value of 188) directs the reader to look at a medicine label to determine the “maximum number of days you should take this medicine.” In terms of our process variables, type of match (TOM) was scored a 1 because the reader was required to locate a single piece of information that was literally stated in the medicine label. The label contained only one reference to number of days and this information was located under the label dosage. Type of information (TOI) was scored a 2 because it asked for a number of days, and plausibility of distractor (POD) received a 1 because there is no other reference to days in the medicine label.

MEDCO ASPIRIN 500

INDICATIONS: Headaches, muscle pains, rheumatic pains, toothaches, earaches. RELIEVES COMMON COLD SYMPTOMS.

DOSAGE: ORAL. 1 or 2 tablets every 6 hours, preferably accompanied by food, for not longer than 7 days. Store in a cool, dry place.

CAUTION: Do not use for gastritis or peptic ulcer. Do not use if taking anticoagulant drugs. Do not use for serious liver illness or bronchial asthma. If taken in large doses and for an extended period, may cause harm to kidneys. Before using this medication for chicken pox or influenza in children, consult with a doctor about Reyes Syndrome, a rare but serious illness. During lactation and pregnancy, consult with a doctor before using this product, especially in the last trimester of pregnancy. If symptoms persist, or in case of an accidental overdose, consult a doctor. Keep out of reach of children.

INGREDIENTS: Each tablet contains
500 mg acetylsalicicylic acid.
Excipient c.b.p. 1 tablet.
Reg. No. 88246

0 67736 11079

Made in Canada by STERLING PRODUCTS, INC.
1600 Industrial Blvd., Montreal, Quebec H9J 3P1

Reprinted by permission

A second prose literacy task directs the reader to look at an article about impatiens. One task receiving a difficulty value of 230 asks the reader: “What happens when the impatiens plant is exposed to temperatures of 14 degrees C or below?” There is a sentence in the text under the section “General Care” that states, “When the plant is exposed to temperatures of 12-14°C, it loses its leaves and won’t bloom anymore.” Like the “Medco” task, this task received a score of 1 for type of match because the reader only needed to make a synonymous match. Unlike the previous task, however, this task received higher scores for type of information and for plausibility of distractor. Type of information was scored 4 because the reader was asked to identify an outcome that occurs when the plant is exposed to certain temperatures. Plausibility of distractor was scored 2 because other numbers are presented in the text and because the previous sentence contains information about the requirements of the plant at other temperatures.

A similar task involving the same text asks the reader to identify “what the smooth leaf and stem suggest about the plant.” This task received a difficulty value of 254. Again, the task directed the reader to locate information contained in the text so it was scored 1 for type of match. The last sentence in the second paragraph under the heading *Appearance* states: “The smooth leaf surfaces and the stems indicate a great need of water.” Type of information was scored a 3 because it directs the reader to identify a condition. Plausibility of distractor was scored a 3 because the same paragraph contained a sentence that serves to distract a number of readers. This sentence states, “… stems are branched and very juicy, which means, because of the tropical origin, that the plant is sensitive to cold.”

IMPATIENS

Like many other cultured plants, impatiens plants have a long history behind them. One of the older varieties was sure to be found on grandmother's windowsill. Nowadays, the hybrids are used in many ways in the house and garden.

Origin: The ancestors of the impatiens, *Impatiens sultani* and *Impatiens holstii*, are probably still to be found in the mountain forests of tropical East Africa and on the islands off the coast, mainly Zanzibar. The cultivated European plant received the name *Impatiens walleriana.*

Appearance: It is a herbaceous bushy plant with a height of 30 to 40 cm. The thick, fleshy stems are branched and very juicy, which means, because of the tropical origin, that the plant is sensitive to cold. The light green or white speckled leaves are pointed, elliptical, and slightly indented on the edges. The smooth leaf surfaces and the stems indicate a great need of water.

Bloom: The flowers, which come in all shades of red, appear plentifully all year long, except for the darkest months. They grow from "suckers" (in the stem's "armpit").

Assortment: Some are compact and low-growing types, about 20 to 25 cm. high, suitable for growing in pots. A variety of hybrids can be grown in pots, window boxes, or flower beds. Older varieties with taller stems add dramatic colour to flower beds.

General care: In summer, a place in the shade without direct sunlight is best; in fall and spring, half-shade is best. When placed in a bright spot during winter, the plant requires temperatures of at least 20°C; in a darker spot, a temperature of 15°C will do. When the plant is exposed to temperatures of 12-14°C, it loses its leaves and won't bloom anymore. In wet ground, the stems will rot.

Watering: The warmer and lighter the plant's location, the more water it needs. Always use water without a lot of minerals. It is not known for sure whether or not the plant needs humid air. In any case, do not spray water directly onto the leaves, which causes stains.

Feeding: Feed weekly during the growing period from March to September.

Repotting: If necessary, repot in the spring or in the summer in light soil with humus (prepacked potting soil). It is better to throw the old plants away and start cultivating new ones.

Propagating: Slip or use seeds. Seeds will germinate in ten days.

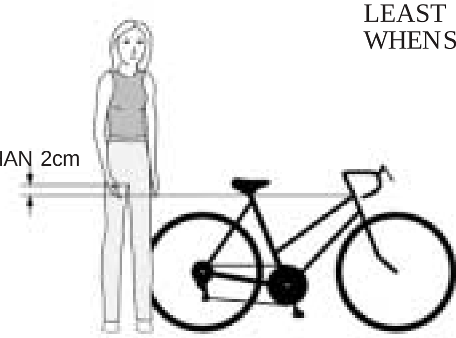
Diseases: In summer, too much sun makes the plant woody. If the air is too dry, small white flies or aphids may appear.

Tasks that fall at higher levels along the scale present the reader with more varied demands in terms of the type of match that is required and in terms of the number and nature of distractors that are present in the text. One such task (with a difficulty value of 281) refers the reader to a page from a bicycle's owner's manual to determine how to ensure the seat is in the proper position. Type of information was scored a 3 because the reader needed to identify and state in writing two conditions that needed to be met. In addition, they were not told how many features they needed to provide from among those stated. Type of information was also scored a 3 because it involved identifying a condition, and plausibility of distractor received a score of 2.

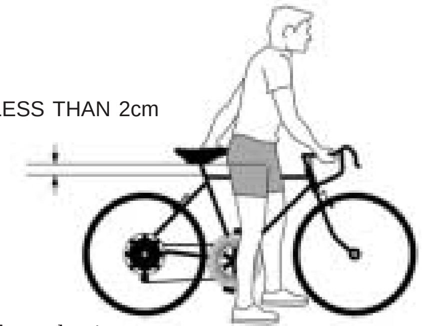
PROPER FRAME FIT

RIDER MUST BE ABLE TO STRADDLE BICYCLE WITH AT LEAST 2 cm CLEARANCE ABOVE THE HORIZONTAL BAR WHEN STANDING.

NOT LESS THAN 2cm



NOT LESS THAN 2cm



NOTE: Measurement for a female should be determined using a men's model as a basis.

OWNER'S RESPONSIBILITY

PROPER SIZE OF BICYCLE	
FRAME SIZE	LEG LENGTH OF RIDER
430mm	660mm-760mm
460mm	690mm-790mm
480mm	710mm-790mm
530mm	760mm-840mm
560mm	790mm-860mm
580mm	810mm-890mm
635mm	860mm-940mm

1. **Bicycle Selection and Purchase:** Make sure this bicycle fits the intended rider. Bicycles come in a variety of sizes. Personal adjustment of seat and handlebars is necessary to assure maximum safety and comfort. Bicycles come with a wide variety of equipment and accessories . . . make sure the rider can operate them.
2. **Assembly:** Carefully follow all assembly instructions. Make sure that all nuts, bolts and screws are securely tightened.
3. **Fitting the Bicycle:** To ride safely and comfortably, the bicycle must fit the rider. Check the seat position, adjusting it up or down so that with the sole of rider's foot on the pedal in its lowest position the rider's knee is slightly bent.

Note: Specific charts illustrated at left detail the proper method of deter-mining the correct frame size.

The manufacturer is not responsible for failure, injury, or damage caused by improper completion of assembly or improper maintenance after shipment.

A somewhat more difficult task (with a difficulty value of 318) involves an article about cotton diapers and directs the reader to "list three reasons why the author prefers to use disposable rather than cotton diapers." This task is made more difficult because of several of our process variables. First, type of match was scored a 5 because the reader had to provide multiple responses, each of which required a text-based inference. Nowhere in the text does the author say, "I prefer cotton diapers because …" These inferences are made somewhat more difficult because the type of information being requested is a "reason" rather than something more concrete. This variable received a score of 4. Finally, plausibility of distractor was scored a 3 because the text contains information that may serve to distract the reader.

An additional task falling at an even higher place along the prose literacy scale (338) directs the reader to use the information from a pamphlet about hiring interviews to "write in your own words one difference between the panel and the group interview." Here the difficulty does not come from locating information in the text. Rather than merely locating a fact about each type of interview, readers need to integrate what they have read to infer a characteristic on which the two types of interviews differ. Experience from other surveys of this kind reveal that tasks in which readers are asked to contrast information are more difficult, on average, than tasks in which they are asked to find similarities. Thus, type of match was scored 6. Type of information was scored 5 because it directs the reader to provide a difference. Differences tend to be more abstract in that they ask for the identification of distinctive or contrastive features related, in this case, to an interview process. Plausibility of distractor was scored 1 because no distracting information was present in the text. Thus, this variable was not seen as contributing to the overall difficulty of this task.

The Hiring Interview

Preinterview

Try to learn more about the business. What products does it manufacture or services does it provide? What methods or procedures does it use? This information can be found in trade directories, chamber of commerce or industrial directories, or at your local employment office.

Find out more about the position. Would you replace someone or is the position newly created? In which departments or shops would you work? Collective agreements describing various standardized positions and duties are available at most local employment offices. You can also contact the appropriate trade union.

The Interview

Ask questions about the position and the business. Answer clearly and accurately all questions put to you. Bring along a note pad as well as your work and training documents.

The Most Common Types of Interview

One-on-one: Self explanatory.

Panel: A number of people ask you questions and then compare notes on your application.

Group: After hearing a presentation with other applicants on the position and duties, you take part in a group discussion.

Postinterview

Note the key points discussed. Compare questions that caused you difficulty with those that allowed you to highlight your strong points. Such a review will help you prepare for future interviews. If you wish, you can talk about it with the placement officer or career counsellor at your local employment office.

CANCO

CANCO Manufacturing Company
Personnel Department

Centre on Internal and External Mobility

What is CIEM?

CIEM stands for Centre on Internal and External Mobility, an initiative of the personnel department. A number of workers of this department work in CIEM, together with members from other departments and outside career consultants.

CIEM is available to help employees in their search for another job inside or outside the Canco Manufacturing Company.

What does CIEM do?

CIEM supports employees who are seriously considering other work through the following activities:

- ***Job Data Bank***

After an interview with the employee, information is entered into a data bank that tracks job seekers and job openings at Canco and at other manufacturing companies.

- ***Guidance***

The employee's potential is explored through career counselling discussions.

- ***Courses***

Courses are being organized (in collaboration with the department for information and training) that will deal with job search and career planning.

- ***Career Change Projects***

CIEM supports and coordinates projects to help employees prepare for new careers and new perspectives.

- ***Mediation***

CIEM acts as a mediator for employees who are threatened with dismissal resulting from reorganization, and assists with finding new positions when necessary.

How much does CIEM cost?

Payment is determined in consultation with the department where you work. A number of services of CIEM are free. You may also be asked to pay, either in money or in time.

How does CIEM work?

CIEM assists employees who are seriously considering another job within or outside the company.

That process begins by submitting an application. A discussion with a personnel counsellor can also be useful. It is obvious that you should talk with the counsellor first about your wishes and the internal possibilities regarding your career. The counsellor is familiar with your abilities and with developments within your unit.

Contact with CIEM in any case is made via the personnel counsellor. He or she handles the application for you, after which you are invited to a discussion with a CIEM representative.

For more information

The personnel department can give you more information.

The most difficult task on the prose literacy scale (377) requires readers to look at an announcement from a personnel department and to "list two ways in which CIEM (an employee support initiative within a company) helps people who lose their jobs because of departmental reorganization." Type of match was scored 7 because the question contained multiple phrases that the reader needed to keep in mind when reading the text. In addition, readers had to provide multiple responses and make low text-based inferences. Type of information was scored 3 because readers were looking for a purpose or function, and plausibility of distractor was scored a 4. This task is made somewhat more difficult because the announcement is organized around information that is different from what is being requested in the question. Thus, while the correct information is listed under a single heading, this information is embedded under a list of headings describing CIEM's activities for employees looking for other work. Thus, the list of headings in this text serves as an excellent set of distractors for the reader who does not search for or locate the phrase in the question containing the conditional information—those who lose their jobs because of a departmental reorganization.

Evaluating the contribution of the variables to task difficulty

The Item Response Theory (IRT) scaling procedures that were used in the IALS constitute a statistical solution to the challenge of establishing one or more scales for a set of tasks with an ordering of difficulty that is essentially the same for everyone. Each scale can be characterized in terms of how tasks are ordered along it. The scale point assigned to each task is the point at which individuals with that proficiency score have a given probability of responding correctly. In IALS, a response probability of 80% (RP80) was used. This means that individuals estimated to have a particular scale score are expected to perform tasks at that point on the scale correctly with an 80% probability. It also means they will have a greater than 80% chance of performing tasks that are lower on the scale. It does not mean, however, that individuals with given proficiencies can never succeed at tasks with higher difficulty values; they may do so some of the time. It does suggest that their probability of success is "relatively" low—that is, the more difficult the task relative to their proficiency, the lower the likelihood of a correct response.

An analogy might help clarify this point. The relationship between task difficulty and individual proficiency is much like the high jump event in track and field, in which an athlete tries to jump over a bar that is placed at increasing heights. Each high jumper has a height at which he or she is proficient—that is, the jumper can clear the bar at that height with a high probability of success, and can clear the bar at lower heights almost every time. When the bar is higher than the athlete's level of proficiency, however, it is expected that the athlete will be unable to clear the bar consistently.

Once the literacy tasks are placed along each of the scales using the criterion of 80% (RP80), it is possible to see to what extent the variables associated with task characteristics explain the placement of tasks along the scales. A multiple regression was run using RP80 as the dependent variable.[3] The independent variables were the three process variables (TOM, TOI, and POD) used to characterize the prose tasks, plus a traditional measure of readability[4] (READ). The results are shown here in Table 1.

Table 1 shows the zero order correlation of each predictor variable with RP80 along with the results of the regression analysis. These data reveal that type of match had the largest zero order correlation with RP80 (.89) and received the largest standardized regression weight, followed by plausibility of distractor and type of information. Together these variables, along with readability, accounted for 89% of the variance in predicting RP80 values.

Table 1

Standardized beta and T-ratios representing the regression of readability and process variables against RP80 values on prose tasks, along with their zero order correlation

Variable	Beta coef.	*T*- ratio	Significance	Corr. w/ RP80
TOM	.74	10.0	.00	.89
TOI	.16	2.3	.03	.55
POD	.20	2.8	.01	.54
READ	.11	1.8	.09	.28

Multiple R = .94
Adjusted R^2 = .87

Easy tasks on the prose literacy scale tended to require readers to make a literal match on the basis of a single piece of concrete information where few, if any, distractors were present in the text. Tasks further along the prose scale become somewhat more varied. While some may still require a single feature match, more distracting information may be present in the text or the match may require a low text-based inference. Some tasks may require the reader to cycle through information to arrive at a correct response. Tasks that are more difficult can take on a variety of characteristics. They may still require the reader to make a match, but usually the reader has to match on multiple features or take conditional information into account. Tasks may also require the reader to integrate information from within a text or to provide multiple responses. The most difficult tasks typically require the reader to make higher-level inferences, process conditional information, and deal with highly plausible distracting information.

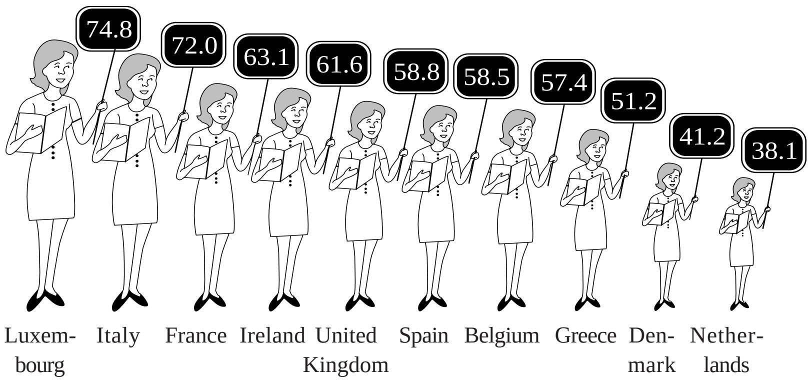
Characterizing document literacy tasks

There are 34 tasks ordered along the IALS 500-point document literacy scale. These tasks range in difficulty value from 182 to 408. One document literacy task with a difficulty value of 188 directs the reader to identify from a chart the percentage of teachers from Greece who are women. The chart shown here displays the percentage of teachers from various countries who are women. In terms of our process variables, type of match (TOM) was scored a 1 because the reader was required to locate a single piece of information that was literally stated in the chart; type of information (TOI) received a 2 because it was an amount; and plausibility of distractor (POD) is also scored a 2 because there are distractors for the requested information.

A second document task involving this same chart directs the reader to identify the country other than the Netherlands in which women teachers are in the minority. This item received a difficulty value of 234. This task was made a bit more difficult than the first because rather than searching on a country and locating a percentage, the readers had to know that minority means less than 50%. Then they had to cycle through to identify the countries in which the percentage of women teachers was less then 50%. In addition, they had to remember the condition “other than the Netherlands”; otherwise, they might have chosen it over the correct response. As a result, type of match was scored a 3; type of information was scored a 1 because the requested information is a country or place; and plausibility of distractor was given a 2 because there are distractors associated with the requested information.

FEW DUTCH WOMEN AT THE BLACKBOARD

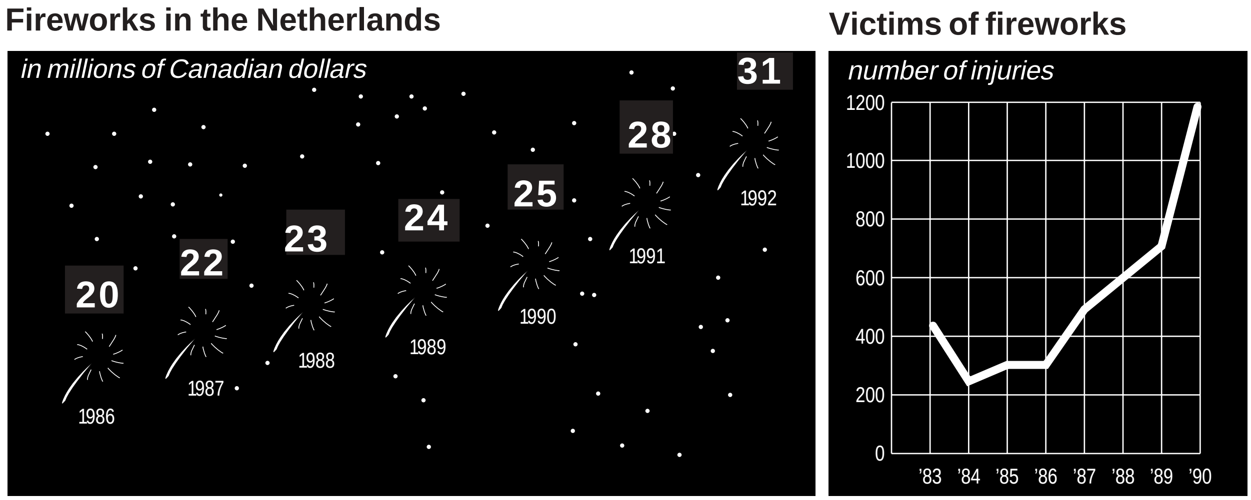
There is a low percentage of women teachers in the Netherlands compared to other countries. In most of the other countries, the majority of teachers are women. However, if we include the figures for inspectors and school principals, the proportion shrinks considerably and women are in a minority everywhere.



Percentage of women teachers (kindergarten, elementary, and secondary).

Another task receiving a difficulty value of 242 is very similar to the one discussed above. This item directs the readers to look at two charts involving fireworks in the Netherlands and to identify the year in which fireworks injured the fewest people. This task also was rated a 3 for type of match because the readers had to first identify which of the two charts contained the requested information. Then they had to cycle through the points of the graph to locate which point represented the fewest injuries. Using this point, they then had to identify the correct year. Type of information received a score of 2 since the requested information was time, and plausibility of distractor received a score of 2 because there were other years the reader could have selected.

A somewhat more difficult task (with a difficulty value of 295) involving the fireworks charts directs the reader to write a brief description of the relationship between sales and injuries based on the information shown. Here the reader needs to look at and compare the information contained in the two charts and integrate this information, making an inference regarding the relationship between the two sets of information. As a result, it was scored a 5 for type of match. Type of information received a 4 because the requested information is asking for a pattern or similarity in the data. Plausibility of distractor was scored 3, primarily because both given and requested information is present in the task. For example, one of the things that may have contributed to the difficulty of this task is the fact that the sales graph goes from 1986 to 1992, while the injuries graph goes from 1983 to 1990. The reader should have compared the information from the two charts for the comparable period of time.



Another set of tasks covering a range of difficulty on the document scale involved a rather complicated document taken from a page in a consumer magazine rating clock radios. The easiest of the three tasks, receiving a difficulty value of 287, asks the reader: "Which two features are not on any basic clock radio?" In looking at the document, the reader has to cycle through the document to find the listing for basic clock radios and then determine that a dash represents the absence of a feature. The reader then has to locate the two features indicated by the set of dashes. As a result, type of match received a score of 4 because it is a cycle requiring multiple responses with a condition or low text-based inference. Type of information was scored a 2 because its features are attributes of the clock radio, and plausibility of distractor is a 2 because there are some characteristics that are not associated with other clock radios.

A somewhat more difficult task associated with this document received a difficulty value of 327 and asks the reader: "Which full-featured clock radio is rated highest on performance?" Here, the reader must make a three-feature match (full-featured, performance, and highest), where one of the features requires the reader to process conditional information. It is possible, for example, that some readers were able to find the full-featured clock radios and the column listed under performance but selected the first clock radio listed, assuming it was the one rated highest. In this case, they did not understand the conditional information, which is a legend stating what the symbols mean. Others may have gone to the column labeled "Overall Score" and found the highest numerical number and chosen the clock radio associated with it. For these reasons, type of match received a score of 4 and plausibility of distractor was scored a 3. Type of information received a 1 because the requested information is a thing.

The most difficult task associated with this document (with a difficulty value of 408) asks the reader to identify the average advertised price for the basic clock radio receiving the highest overall score. This task was made more difficult because the reader had to match four rather than three features; the reader also had to process conditional information, and there was a highly plausible distractor in the same node as the correct answer. As a result of these factors, type of match received a score of 5, type of information a score of 2, and plausibility of distractor a score of 5.

RATINGS

Better ⊙ ◓ ○ ◒ ● Worse

Clock radios

Listed by types; within types, listed in order of overall score. Differences in score of 4 points or less were not deemed significant.

1 Brand and model. If you can't find a model, call the company. Phone numbers are listed on page 736.

2 Price. The manufacturer's suggested or approximate retail price, followed by the average advertised price.

3 Dimensions. To the nearest centimetre.

4 Overall score. A composite, encompassing all our tests and judgments. A "perfect" radio would have earned 100 points.

5 Convenience. This composite judgment reflects such things as the legibility of the display, the ease of tuning the radio and setting the alarm, and the presence or absence of useful features.

6 Performance. An overall judgment reflecting performance in our tests of: sensitivity and selectivity; tuning ease; capture ratio, the ability to bring in the stronger of two stations on the same frequency; image rejection, the ability to ignore signals from just above the band, resistance to interference from signals bouncing off aircraft and such.

7 Sensitivity. How well each radio received a station with little interference.

8 Selectivity. How well each radio received clearly a weak station next to a strong one on the dial.

9 Tone quality. Based mainly on computer analysis of the speaker's output and on listening tests, using music from CDs. No model produced high-fidelity sound.

10 Reversible time-setting. This useful feature makes setting clock and alarm times easy. If you overshoot the desired setting, you simply back up.

11 Dual alarm. Lets you set two separate wake-up times.

1 Brand and model	2 Price	3 Dimensions HxWxD, cm	4 Overall Score	5 Convenience	6 Performance	7 Sensitivity	8 Selectivity	9 Tone quality	10 Reversible time setting	11 Dual alarm	Warranty, months	Advantages	Disadvantages	Comments
Full-featured clock radios														
RCA RP-3690	$50/$40	8x25x18	86	⊙	◓	◓	◓	⊙	✓	✓	12	A,B,D,H,J,L,O,T,U		*A*
Sony ICF-C303	50/45	5x20x15	84	⊙	◓	⊙	○	◒	✓	✓	12	C,E,F,I,N,T		*C*
Panasonic RC-X220	50/45	10x28x13	82	◓	◓	◓	◓	○	✓	✓	12	A,G,K,M,O,S,T,U	b,c	*A*
Realistic 272	50/30	5x28x15	79	◓	○	◓	◓	◒	✓	✓	3	A,G,H,K,O,T		*D*
Magnavox AJ3900	65/—	15x38x13	78	○	⊙	◓	⊙	◓	—	✓	3	D,G,K,M,O,R,T	b,g	*B*
Emerson AK2745	39/20	8x28x15	70	○	◒	◒	◒	○	✓	✓	3	G,O	g	*K*
Soundesign 3753	20/20	8x23x13	62	○	◒	●	○	○	✓	✓	3	J,Q	d,h	*J*
Basic clock radios														
Realistic 263	28/18	10x20x10	74	○	◓	◓	○	◒	—	—	3	A,D,H,O,P,U	h	—
Soundesign 3622	12/10	5x20x13	68	◒	◓	⊙	◒	◒	—	—	3	U	d	*L*
Panasonic RC-6064	18/15	5x20x13	67	◒	◓	⊙	○	○	—	—	12	—	b,c	—
General Electric 7-4612	13/10	5x20x13	66	◒	○	⊙	◒	○	—	—	12	A,D	a,g	—
Lloyds CR001	20/15	5x18x13	64	◒	○	⊙	○	◒	—	—	3	U	—	—
Sony ICF-C240	15/13	5x18x15	63	◒	○	○	○	○	—	—	12	—	f,g	—
Emerson AK2720	19/10	5x20x13	61	◒	○	◓	●	○	—	—	3	O,T	e	*K*
Gran Prix D507	15/10	5x18x10	54	◒	●	○	●	●	—	—	3	—	d	—
Clock radios with cassette player														
General Electric 7-4965	60/50	10x30x15	85	◓	◓	◓	◓	◓	✓	✓	12	A,D,G,H.K,O,S,T	—	*B,E*
Panasonic RC-X250	[1]	10x33x13	76	◓	◓	○	◓	◓	✓	✓	12	A,G,K,O,R,U	b,c	*A,H*
Sony ICF-CS650	75/65	15x28x15	74	○	◓	⊙	○	◒	✓	✓	12	G,R,T,U	c,f,i	A,F,H
Soundesign 3844MGY	40/30	13x30x13	62	○	●	●	●	◒	—	—	3	G,K,J,S,U		*F,G,I,M*

[1] *Discontinued. Replaced by* **RC-X260,** *$79 list and $60 average advertised sale price.*

Features in Common

All: • Permit snooze time of about 8 min. • Retain time settings during short power failures.

Except as noted, all have: • Battery backup for clock and alarm memory. • Red display digits 1 cm. high. • Sleep-time radio play for up to 60 min. before automatic shutoff. • Switch to reset alarm.

Keys to Advantages

A – Alarm works despite power failure.
B – Shows actual time plus up to 2 alarm times.
C – Twin alarms settable for 2 different stations.
D – Tone alarm has adjustable volume control.
E – Memory needs no battery.
F – Digital tuner with presettable stations.
G – Tuner can receive in stereo.
H – Battery-strength indicator.
I – Illuminated tuning dial.
J – Illuminated tuning pointer.
K – Earphone jack.
L – Nap timer.
M – Audio input for tape deck or CD player.
N – Display can show date and time.
O – Display has high/low brightness switch.
P – Display has larger digits than most.
Q – Night light—adjusts for room light.
R – Bass-boost tone control.
S – Treble-cut tone control.
T – Better than most in tuning ease.
U – Better than most in image rejection.

Key to Disadvantages

a – Possible to reset time by accident.
b – Controls for time-setting or dimmer inconveniently located on radio's bottom or rear.
c – Display dimmer than most in brightly lit room.
d – Radio volume must be turned completely down for alarm buzzer to sound.
e – Lacks alarm buzzer; radio is sole alarm.
f – Lacks indication alarm is set.
g – Lacks alarm-reset button.
h – Time-setting lacks fast reverse.
i – No slow forward, fast reverse for time setting.

Key to Comments

A – Display shows green digits.
B – Display shows blue digits.
C – Display uses LCD (liquid crystal) digits.
D – Terminals for external antenna.
E – 3-position graphic equalizer.
F – Cassette player lacks Record function.
G – Cassette player lacks Rewind function.
H – Model permits wake-up to cassette play.
I – Cassette-deck flutter worse than most.
J – Warranty repairs cost $3 for handling.
K – Warranty repairs cost $3.50 for handling.
L – Warranty repairs cost $6 for handling.
M – Warranty repairs cost $10 for handling.

Evaluating the contribution of the variables to task difficulty

As with the prose scale, IRT was used to establish the document literacy scale as well as to characterize tasks along it. Again, a response probability of 80% was used as an indicator that someone at a specified point on the document literacy scale has mastered or is proficient with tasks at that place on the scale. It does not mean that they cannot perform tasks above their estimated proficiency; rather, they may do so, but with less consistency. Their expected consistency on tasks above their level of proficiency depends on how far the task is from their estimated proficiency.

Once the document literacy tasks are placed along each of the scales using the criterion of 80% (RP80), it is possible to determine to what extent the variables associated with the task characteristics explain the placement of tasks along the scales. A multiple regression was run using RP80 as the dependent variable (see note number 3). The independent variables were the three process variables (TOM, TOI, and POD) used to characterize the prose and document literacy tasks, plus a newly developed measure of document readability (READ) (Mosenthal and Kirsch, 1998).[5] The results are shown here in Table 2.

Table 2

Standardized beta and T-ratios representing the regression of readability and process variables against RP80 values on document tasks, along with their zero order correlation

Variable	Beta coef.	*T*- ratio	Significance	Corr. w / RP80
TOM	.43	3.7	.00	.85
TOI	.13	1.4	.16	.43
POD	.40	3.8	.00	.71
READ	.17	1.7	.09	.55

Multiple R = .89
Adjusted R^2 = .76

Table 2 shows the zero order correlation between each of the predictor variables and RP80, along with the results from the regression analysis. These data reveal that each of the predictor variables is significantly correlated with RP80, yet only two process variables received significant beta weights. It should be noted that while each of these variables may not be significant in terms of this regression analysis, each was taken into consideration when constructing the literacy tasks and, therefore, each is important as to how well the domain is represented. Together the set of variables accounted for 79% of the variance in RP80 values. Type of match received the largest standardized regression weight, followed by plausibility of distractors.

Easy tasks on the document literacy scale tended to require readers to make a literal match on the basis of a single piece of information. Tasks further along the document scale become somewhat more varied. While some may still require a single feature match, more distracting information may be present in the document or the match may require a low text-based inference. Some tasks may require the reader to cycle through information to arrive at a correct response. Tasks that are more difficult can take on a variety of characteristics. They may still require the reader to make a match, but usually the reader has to match on multiple features or take conditional information into account. Tasks may also require the reader to integrate information from one or more documents, or cycle through a document to provide multiple responses. The most difficult tasks typically require the reader to match on multiple features, to cycle through documents, and to integrate information. Frequently, these tasks require the reader to make higher-level inferences, process conditional information, and deal with highly plausible distractors. These tasks also tend to be associated with more complex displays of information.

Characterizing quantitative literacy tasks

There are 33 tasks ordered along the IALS 500-point quantitative literacy scale. These tasks range in difficulty value from 225 to 409. The easiest quantitative literacy task (with a difficulty value of 225) directs the reader to complete an order form. The last line on this form says, "Total with Handling." The line above it says, "Handling Charge $2.00." The reader simply had to add the $2.00 to the $50.00 they had entered on a previous line to indicate the cost of the tickets. In terms of our process variables, this item received a code of 1. The design of the form set the problem up in simple column format for the reader and the amount for handling was stipulated, so there was little required of them in terms of type of match (TOM) or plausibility of distractor (POD). In addition, the last line on the form said, "Total with Handling," indicating the type of operation and the numbers did not require the reader to carry or borrow. As a result, both type of calculation (TOC) and operation specificity (OSP) were each coded 1.

A second quantitative literacy task directs the reader to use a weather chart in a newspaper to determine how many degrees warmer today's high temperature is expected to be in Bangkok than in Seoul. This item received a difficulty value of 255. This task was made more difficult both in terms of the literacy processes and in terms of those processes associated with the quantitative scale. Here the reader had to cycle through a complex table to make two, three-feature matches to identify the two temperatures, and then subtract one from the other to determine the difference. The numbers they had to subtract were not adjacent to each other in the table, were not in column format, and had to be identified through a search. As a result, operation specificity was coded a 3, type of calculation received a 2, type of match was scored a 4, and plausibility of distractor was scored a 4.

WEATHER

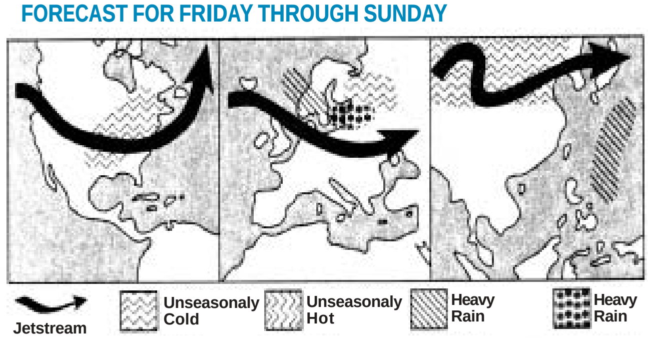
Europe

	Today High C	Today Low C	W	Tomorrow High C	Tomorrow Low C	W
Algarve	19	7	s	21	9	s
Amsterdam	11	6	pc	12	7	pc
Ankara	17	7	pc	19	8	pc
Athens	22	15	pc	23	14	pc
Barcelona	16	8	s	14	9	s
Belgrade	14	6	pc	10	1	c
Berlin	8	2	c	6	1	c
Brussels	11	6	pc	14	7	pc
Budapest	9	1	pc	9	2	c
Copenhagen	7	1	r	6	2	c
Costa del Sol	21	8	s	21	10	s
Dublin	10	6	pc	13	8	pc
Edinburgh	10	6	c	10	6	c
Florence	11	5	s	14	6	s
Frankfurt	12	6	pc	13	4	pc
Geneva	9	2	s	12	4	s
Helsinki	-1	-7	sf	-3	-10	pc
Istanbul	17	10	pc	15	9	sh
Las Palmas	26	18	pc	27	18	pc
Lisbon	19	9	s	19	10	s
London	12	5	pc	13	7	pc
Madrid	17	3	s	18	4	s
Milan	9	3	s	13	6	s
Moscow	1	-3	r	-3	-11	sf
Munich	11	3	pc	12	6	pc
Nice	14	7	s	15	8	s
Oslo	4	-4	c	5	-2	c
Paris	12	6	pc	13	6	pc
Prague	11	1	pc	8	2	c
Reykjavik	4	2	r	6	-1	c
Rome	20	12	s	20	10	s
St. Petersburg	-1	-7	sf	-4	-12	pc
Stockholm	1	-5	sn	-2	-7	c
Strasbourg	12	5	pc	15	7	pc
Tallinn	-1	-7	sf	-4	-10	pc
Venice	10	3	s	11	4	s
Vienna	9	-1	pc	10	2	c
Warsaw	8	2	sh	6	1	c
Zurich	8	0	s	9	1	pc

Oceania

Auckland	20	14	s	17	11	sh
Sydney	27	17	pc	25	16	pc

FORECAST FOR FRIDAY THROUGH SUNDAY



North America
Cold weather will engulf the Midwestern and Northeastern United States Friday and over the weekend. Although it will be cold in Chicago, Toronto and New York City, the weather is expected to be dry. Los Angeles will have some sunshine and seasonable temperatures each day.

Europe
Western and central Europe will have a spell of mild weather Friday into the weekend. London and Paris will have dry weather with some sunshine Friday into Sunday. Rain will continue to soak southwestern Norway. Snow will blanket the area from Minsk to Moscow.

Asia
Typhoon Elsie will probably stay to the east of the Philippines and south of Japan Friday and the weekend. Some rain is apt to fall in Seoul and there could even be a little ice or snow. Cold air will pour into Beijing and snow is a possibility. Hong Kong will start the weekend warm.

Middle East

	Today High C	Today Low C	W	Tomorrow High C	Tomorrow Low C	w
Beirut	28	19	pc	29	20	s
Cairo	29	20	pc	28	19	pc
Damascus	24	12	s	26	14	s
Jerusalem	27	15	s	26	14	s
Riyadh	34	13	s	32	13	s

Latin America

	Today High C	Today Low C	W	Tomorrow High C	Tomorrow Low C	W
Buenos Aires	23	11	pc	26	13	s
Caracas	29	20	s	31	18	s
Lima	23	17	c	23	16	c
Mexico City	23	11	sh	23	12	pc
Rio de Janiero	32	22	s	28	21	sh
Santiago	24	4	s	22	6	pc

Legend: s-sunny, pc-partly cloudy, c-cloudy, sh-showers, t-thunderstorms, r-rain, sf-snow flurries, sn-snow, i-ice, W-Weather. **All maps, forecasts and data provided by Accu-Weather, Inc. © 1992**

Asia

	Today High C	Today Low C	W	Tomorrow High C	Tomorrow Low C	W
Bangkok	32	22	pc	30	23	s
Beijing	11	0	s	8	2	pc
Hong Kong	30	23	s	29	22	pc
Manila	31	25	s	31	25	sh
New Delhi	31	13	s	32	16	s
Seoul	14	6	pc	14	4	pc
Shanghai	22	10	pc	24	12	s
Singapore	31	24	pc	28	23	sh
Taipei	26	21	pc	26	19	pc
Tokyo	18	9	pc	17	7	pc

Africa

Algiers	27	14	s	26	13	s
Cape Town	20	11	sh	18	11	pc
Casablanca	20	14	c	21	11	pc
Harare	34	17	s	32	18	pc
Lagos	30	24	pc	29	24	pc
Nairobi	27	12	pc	26	13	pc
Tunis	27	17	pc	17	14	pc

North

Anchorage	0	-2	c	3	0	sh
Atlanta	14	4	pc	8	2	pc
Boston	15	4	c	8	-1	pc
Chicago	2	-5	c	-2	-8	pc
Denver	8	-3	pc	4	-6	sn
Detroit	4	-2	c	4	-5	pc
Honolulu	31	20	s	31	21	pc
Houston	15	3	pc	12	6	pc
Los Angeles	28	14	s	24	13	s
Miami	30	22	pc	29	21	pc
Minneapolis	-1	-8	c	1	-7	pc
Montreal	7	-2	sf	4	-3	c
Nassau	31	22	pc	28	21	sh
New York	14	4	r	10	2	pc
Phoenix	23	11	pc	22	8	s
San Fran.	20	11	pc	21	8	s
Seattle	11	6	pc	13	7	r
Toronto	6	-3	c	3	-3	c
Washington	14	6	r	11	4	pc

A similar but slightly more difficult task (with a difficulty value of 268) requires the reader to use the chart about women in the teaching profession that is displayed under the document scale. This task directs the reader to calculate the percentage of men in the teaching profession in Italy. Both this task and the one just described above involve calculating the difference between two numbers. In the former, however, both numbers could be located by matching on information displayed in the table of temperatures taken from the newspaper. For the task involving male teachers in Italy, the reader must make the inference that percentage of male teachers is equal to 100% minus the percentage of female teachers. Thus, while type of calculation, type of match, and plausibility of distractor each received a code of 2, operation specificity was coded a 5, suggesting that this might be a slightly more difficult task in terms of this variable.

Tasks falling around 300 on the quantitative scale still require the reader to perform a single arithmetic operation, but the quantities may not be as easily determined. For example, one task, located at 293 on the quantitative scale, directs the reader to look at the chart depicting fireworks shown earlier for documents. The question directs the reader to calculate how many more people were injured in 1989 than in 1988. As with the earlier tasks, this task also requires the reader to subtract the difference between two quantities. Part of what contributes to the increased difficulty of this task is the fact that the reader first must determine which of the two charts is the correct one to use for this task. In addition, one of the numbers needed is not stated in the graph but must be interpolated from the information provided along the vertical axis. As a result, type of match was scored 4, plausibility of distractor was scored 2, type of calculation was scored 2, and operation specificity was coded 5.

Compound Interest
Compounded Annually

Principal	**Period**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**	**12%**	**14%**	**16%**
$100	1 day	0.011	0.014	0.016	0.019	0.022	0.025	0.027	0.033	0.038	0.044
	1 week	0.077	0.096	0.115	0.134	0.153	0.173	0.192	0.230	0.268	0.307
	6 mos	2.00	2.50	3.00	3.50	4.00	4.50	5.00	6.00	7.00	8.00
	1 year	4.00	5.00	6.00	7.00	8.00	9.00	10.00	12.00	14.00	16.00
	2 years	8.16	10.25	12.36	14.49	16.64	18.81	21.00	25.44	29.96	34.56
	3 years	12.49	15.76	19.10	22.50	25.97	29.50	33.10	40.49	48.15	56.09
	4 years	16.99	21.55	26.25	31.08	36.05	41.16	46.41	57.35	68.90	81.06
	5 years	21.67	27.63	33.82	40.26	46.93	53.86	61.05	76.23	92.54	110.03
	6 years	26.53	34.01	41.85	50.07	58.69	67.71	77.16	97.38	119.50	143.64
	7 years	31.59	40.71	50.36	60.58	71.38	82.80	94.87	121.07	150.23	182.62
	8 years	36.86	47.75	59.38	71.82	85.09	99.26	114.36	147.60	185.26	227.84
	9 years	42.33	55.13	68.95	83.85	99.90	117.19	135.79	177.31	225.19	280.30
	10 years	48.02	62.89	79.08	96.72	115.89	136.74	159.37	210.58	270.72	341.14
	12 years	60.10	79.59	101.22	125.22	151.82	181.27	213.84	289.60	381.79	493.60
	15 years	80.09	107.89	139.66	175.90	217.22	264.25	317.72	447.36	613.79	826.55
	20 years	119.11	165.33	220.71	286.97	366.10	460.44	572.75	864.63	1,274.35	1,846.08

More difficult tasks on the quantitative scale require readers to perform an arithmetic operation where the quantities and/or the operation are not easily determined. One such task involves a compound interest table. It directs the reader to "calculate the total amount of money you will have if you invest $100 at a rate of 6% for 10 years." This task received a difficulty value of 348, in part because many respondents treated it as a document rather than a quantitative task and simply looked up the amount of interest that would be earned. They forgot to add it to the initial investment of $100. Clearly, it was not the arithmetic of adding these two values together that increased difficulty. Rather, it was locating the correct amount of interest in the table and then knowing or inferring that it had to be added to the initial investment stated in the directive. As a result, operation specificity received a code of 6, type of match was scored 2, plausibility of distractor was scored 3, and type of calculation was scored 1 because the reader had only to add to decimal numbers.

Another task at this level requires respondents to read a newspaper article describing a research finding linking allergies to a particular genetic mutation. The question directs the reader to calculate the number of people studied who were found to have a mutant gene. To answer the question correctly the respondent had to know how to set up the problem with the information given. That is, they had to know they needed to convert the phrase "64 percent" to a decimal number and then multiply it by the number of persons studied. The short newspaper article provided no clues on how to set up this problem. As a result, type of calculation was coded 3 because it involved a multiplication, and operation specificity was coded 6 because it required the reader to convert to a decimal and to infer the operation that was needed. Type of match and plausibility of distractor each received a code of 1.

One of the most difficult quantitative literacy tasks directs the reader to look at a table providing nutritional analysis of food and then, using the information given, determine the percentage of calories in a Big Mac® that comes from total fat. This task was at 381 on the scale as a result of how readers responded to this task. To answer this question, readers first must cycle through a long table with lots of distractors to identify the correct numbers needed for this task. Next, they must recognize that the information about total fat is provided in grams. Therefore, they must convert the number of fat grams to calories before calculating this number of calories as a percentage of the total calories given for a Big Mac®. As a result, type of match and plausibility of distractor each received a code of 4. Type of calculation was scored a 5 because the task required multiple calculations, and operation specificity received a score of 9 because of the inferencing needed to discern the features of the problem and to set it up correctly.

Nutritional Analysis

	Serving Size	Calories	Protein (g)	Carbohydrates (g)	Total Fat (g)	Saturated Fat (g)	Monounsaturated Fat (g)	Polyunsaturated Fat (g)	Cholesterol (mg)	Sodium (mg)
Sandwiches										
Hamburger	102 g	255	12	30	9	5	1	3	37	490
Cheeseburger	116 g	305	15	30	13	7	1	5	50	725
Quarter Pounder®	166 g	410	23	34	20	11	1	8	85	645
Quarter Pounder® w/Cheese	194 g	510	28	34	28	16	1	11	115	1110
McLean Deluxe™	206 g	320	22	35	10	5	1	4	60	670
McLean Deluxe™ w/Cheese	219 g	370	24	35	14	8	1	5	75	890
Big Mac®	215 g	500	25	42	26	16	1	9	100	890
Filet-O-Fish®	141 g	370	14	38	18	8	6	4	50	730
McChicken®	187 g	415	19	39	19	9	7	4	50	830
French Fries										
Small French Fries	68 g	220	3	26	12	8	1	2.5	0	110
Medium French Fries	97 g	320	4	36	17	12	1.5	3.5	0	150
Large French Fries	122 g	400	6	46	22	15	2	5	0	200
Salads										
Chef Salad	265 g	170	17	8	9	4	1	4	111	400
Garden Salad	189 g	50	4	6	2	1	0.4	0.6	65	70
Chunky Chicken Salad	255 g	150	25	7	4	2	1	1	78	230
Side Salad	106 g	30	2	4	1	0.5	0.2	0.3	33	35
Croutons	11 g	50	1	7	2	1.3	0.1	0.5	0	140
Bacon Bits	3 g	15	1	0	1	0.3	0.2	0.5	1	95

Soft Drinks

	Coca-Cola Classic®				diet Coke®				Sprite®			
	Small	Medium	Large	Jumbo	Small	Medium	Large	Jumbo	Small	Medium	Large	Jumbo
Calories	140	190	260	380	1	1	2	3	140	190	260	380
Carbohydrates (g)	38	50	70	101	0.3	0.4	0.5	0.6	36	48	66	96
Sodium (mg)	15	20	25	40	30	40	60	80	15	20	25	40

Evaluating the contribution of the variables to task difficulty

As with the prose and document scales, IRT was used to establish the quantitative literacy scale as well as to characterize tasks along it. Again, a response probability of 80% was used as an indicator that someone at a specified point on the quantitative literacy scale has mastered or is proficient with tasks at that place on the scale. It does not mean that they cannot perform tasks above their estimated proficiency; rather, they may do so, but with less consistency. Their expected consistency on tasks above their level of proficiency depends on how far the task is from their estimated proficiency.

Once the quantitative literacy tasks are placed along the scale using the criterion of 80% (RP80), it is possible to determine to what extent the variables associated with task characteristics explain the placement of tasks along the scales. A multiple regression was run using RP80 as the dependent variable (see note 3). The independent variables were the two process variables used to characterize the prose and document literacy tasks—type of match (TOM) and plausibility of distractor (POD)—plus a newly developed measure of document readability (READ) (Mosenthal and Kirsch, 1998).[6] Type of information (TOI) is a constant on this scale since each question requires the reader to determine an amount. In addition, we included the two process variables created for the quantitative scale—type of calculation (TOC) and operation specificity (OSP). The results are shown here in Table 3.

Table 3 shows the zero order correlation between each of the predictor variables and RP80, along with output from the regression analysis. These data reveal that operation specificity, type of calculation, and plausibility of distractor had the highest zero order correlation with RP80. In terms of the regression analysis, operation specificity received the largest standardized regression weight, followed by plausibility of distractor. Neither readability nor the other process variables were significant predictors on this set of tasks. As with the prose and document scales, it is important to note that while only some of these variables receive significant weights in the model, each is important in constructing the quantitative literacy tasks and in representing the domain. Together this set of variables accounted for 75% of the variance in RP80 values.

Table 3

Standardized beta and T-ratios representing the regression of readability and process variables against RP80 values on quantitative tasks, along with their zero order correlation

Variable	Beta coef.	*T*- ratio	Significance	Corr. w / RP80
OSP	.64	5.2	.00	.78
TOC	.16	1.4	.18	.54
TOM	-.18	-1.5	.14	.26
POD	.40	3.3	.00	.50
READ	.05	.4	.09	.33

Multiple R = .87
Adjusted R^2 = .70

Easy tasks on the quantitative literacy scale tended to require readers to perform a single, relatively simple operation (addition), where either the numbers are clearly noted or provided in the text or and the operation is stipulated. Slightly more difficult tasks may require the reader to perform either an addition or a subtraction with numbers that are relatively easy to locate in the text but where the operation can be easily inferred from the wording in the question or directive. Tasks further along the quantitative scale become more varied both in terms of the type of operation they may be asked to perform, and in terms of the extent to which the numbers are embedded in more complex displays or the amount of inferencing that may be required to determine the appropriate operation that is needed. A distinguishing characteristic of the most difficult tasks along this scale is the fact that the reader is required to perform multiple operations sequentially and they must discern the features of the problem from the material and directive given.

Building an interpretative scheme

Identifying and validating a set of variables that predict performance along each of the literacy scales provides a basis for building an interpretative scheme. This scheme provides a useful means for exploring the progression of information-processing demands across each of the scales and what scores along a particular scale mean. Thus, it contributes to the construct validity of inferences based on scores from a measure (Messick, 1989). This section summarizes an interpretative scheme that was adopted by IALS. The procedure builds on Beaton's anchored proficiency procedures (Beaton and Allen, 1992; Messick, Beaton, and Lord, 1983), but it is more flexible and inclusive than the one originally developed and used in the 1980s by NAEP. It has been used in various large-scale surveys of literacy in North America (Kirsch and Jungeblut, 1992; Kirsch et al., 1993).

As shown in the previous section of this paper, there is empirical evidence that a set of variables can be identified that summarize some of the skills and strategies that are involved in accomplishing various kinds of prose, document, and quantitative literacy tasks. More difficult tasks tend to feature more varied and complex information-processing demands than are required by easier tasks. This suggests that literacy is neither a single skill suited to all types of tasks nor an infinite number of skills each associated with a particular type of task.

In the North American literacy surveys, when researchers coded each literacy task in terms of the process variables described in this paper they noted that the values for these variables tended to "shift" at various places along each of the literacy scales. These places seemed to be around 50-point intervals, beginning around 225 on each scale (Kirsch et al., 1998). While most of the tasks at the lower end of the scales had code values of 1 on each of the process variables, tasks with scores around 225 were more likely to have code values of 2. Among tasks with scores around 275, many of the codes were 2s and an increasing number were 3s. Among tasks with response probability values of 325, at least one of the three variables had a code value of 4. Code values of 4 or higher predominated tasks at around 375 or higher on the literacy scales.

Although there were some variations across the literacy scales at the points where the coding shifts occurred, the patterns were remarkably consistent. Further, as was shown in this paper with the IALS tasks, this system of coding tasks accounts for much (although not all) of the variance associated with tasks along the literacy scales. Based on these findings, researchers defined five levels of proficiency having the following score ranges:

Level 1: 0–225

Level 2: 226–275

Level 3: 276–325

Level 4: 326–375

Level 5: 376–500

Once the literacy levels were identified based on the noted shifts in code values for the three process variables, criteria were identified that would describe the placement of tasks within these levels. These criteria are summarized along with the data to which they were applied in a chapter appearing in the IALS technical report (Kirsch et al., 1998). Based on evidence resulting from this work, the five literacy levels were used for reporting results from literacy surveys in both national and international surveys using these literacy scales.

One of the advantages resulting from this approach for reporting results is the fact that it is possible to estimate the probability that an individual who is estimated to be in a particular literacy level will be able to perform the typical task in that level and in other levels. Unlike traditional test scores, which provide a single estimate of ability, these probability estimates offer a richer and more accurate reflection of the range of tasks that a person can be expected to perform successfully. After all, while each individual task used in an assessment is of some interest and importance, we are more likely to be interested in the class of tasks each item is intended to represent—that is, items that have similar characteristics and that we want to generalize outside the testing situation. Any assessment is likely to be more useful if we are able to generalize from the particular items used in the survey to the set of behaviors we are most concerned about.

These results mean that the literacy levels not only provide a means for exploring the progression of information-processing demands across each of the literacy scales, but they also can be used to help explain how the proficiencies demonstrated by various countries and various subpopulations reflect the likelihood they will respond correctly to a broad range of tasks used not only in IALS but to tasks having similar characteristics as well. In practical terms, this means that individuals performing at 250 on a literacy scale are expected to be able to perform the average Level 1 and Level 2 task with a high degree of proficiency. That is, they are expected to be able to perform these kinds of tasks with an average probability of 80% or higher. It is important to note that this does not mean they will not be able to perform correctly on literacy tasks in Levels 3 or higher. They will be expected to do so some of the time, but not with the same level of consistency.

The three tables shown here (Tables 4, 5, and 6) display the probability that individuals performing at selected points on each of the scales will give a correct response to tasks of varying difficulty. For example, Table 4 shows that a reader whose prose proficiency is 150 has less than a 50% chance of giving a correct response to the Level 1 tasks. Individuals whose proficiency score is 200, in contrast, have about an 80% probability of responding correctly to these tasks.

In terms of task demands, it can be inferred that adults performing at 200 on the prose scale are likely to be able to locate a single piece of information in a brief text when there is no distracting information, or if plausible but incorrect information is present but located away from the correct answer. However, these individuals are likely to encounter far more difficulty with tasks in Levels 2 through 5. For example, they would have only a 40% chance of performing the average Level 2 task correctly, an 18% chance of success with tasks in Level 3, and no more than a 7% chance with tasks in Levels 4 and 5.

In contrast, respondents demonstrating a proficiency of 300 on the prose scale have about an 80% chance or higher of succeeding with tasks in Levels 1, 2, and 3. This means that they demonstrate success with tasks that require them to make low level inferences and with those that entail taking some conditional information into account. They can also integrate or compare and contrast information that is easily identified in the text. On the other hand, they are likely to encounter difficulty with tasks where they must make more sophisticated textbased inferences, or where they need to process more abstract types of information. These more difficult tasks may also require them to draw on less familiar or more specialized types of knowledge beyond that given in the text. On average, they have about a 50% probability of performing Level 4 tasks correctly; with Level 5 tasks, their likelihood of responding correctly decreases to 40%.

Table 4
Average probabilities of successful performance, prose scale

			Selected proficiency scores		
Prose level	150	200	250	300	350
			%		
1	48	81	95	99	100
2	14	40	76	94	99
3	6	18	46	78	93
4	2	7	21	50	80
5*	2	6	18	40	68

* Probabilities are based on one task.

Table 5
Average probabilities of successful performance, document scale

			Selected proficiency scores		
Document level	150	200	250	300	350
			%		
1	40	72	94	99	100
2	20	51	82	95	99
3	7	21	50	80	94
4*	4	13	34	64	85
5*	<1	1	3	13	41

* Probabilities are based on one task.

Table 6
Average probabilities of successful performance, quantitative scale

			Selected proficiency scores		
Quantitative level	150	200	250	300	350
			%		
1*	34	67	89	97	99
2	20	45	75	92	98
3	7	20	48	78	93
4	1	6	22	58	87
5	<1	2	7	20	53

* Probabilities are based on one task.

Similar kinds of interpretations can be made using the information presented for the document and quantitative scales. For example, someone who is at 250 on the document scale has, on average, an 82% chance of responding correctly to Level 2 tasks. His or her likelihood of responding correctly decreases to 50% for Level 3 tasks, 34% for Level 4 tasks, and only 3% for Level 5 tasks. Similarly, someone at 300 on the quantitative scale has a 78% chance of responding correctly to tasks at this level, but only a 58% chance with Level 4 tasks and a 20% chance with Level 5 tasks. Conversely, they would be expected to perform Level 1 and 2 tasks correctly more than 90% of the time.

Conclusion

One of the goals of large-scale surveys is to provide information that can help policymakers during the decision-making process. Presenting that information in a way that will enhance understanding of what has been measured and the conclusions to be drawn from the data is important to reaching this goal. This paper offers a framework that has been used for both developing the tasks used to measure literacy as well as for understanding the meaning of what is being reported with respect to the comparative literacy proficiencies of adults. The framework identifies a set of variables that have been shown to underlie successful performance on a broad array of literacy tasks. Collectively, they provide a means for moving away from interpreting survey results in terms of discrete tasks or a single number, and toward identifying levels of performance sufficiently generalized to have validity across assessments and groups. As concern ceases to center on discrete behaviors or isolated observations and focuses more on providing meaningful interpretations of performance, a higher level of measurement is reached (Messick, 1989).

References

Almond, R.G., and Mislevy, R.J. (1998). *Graphical models and computerized adaptive testing*. (TOEFL Tech. Rep. No. 14). Princeton, NJ: Educational Testing Service.

Beach, R., and Appleman, D. (1984). Reading strategies for expository and literacy text types. In A. Purves and O. Niles (Eds.), *Becoming readers in a complex society: Eighty-third yearbook of the National Society for the Study of Education* (p.115). Chicago: University of Chicago Press.

Beaton, A.E., and Allen, N.L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics, 17,* 191–204.

Cattel, R.B. (1966). The screen test for the number of factors. *Multivariate Behavioral Research*, *1*, 140–161.

Clark, H., and Haviland, S.E. (1977). Comprehension and the given–new contract. In R.O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–39). Norwood, NJ: Ablex.

Cook-Gumperz, J., and Gumperz, J. (1981). From oral to written culture: The transition to literacy. In M. Whitman (Ed.), *Writing: The nature, development and teaching of written communication: Vol. 1*. Hillsdale, NJ: Erlbaum.

Crandall, J. (1981, December). *Functional literacy of clerical workers: Strategies for minimizing literacy demands and maximizing available information*. Paper presented at the annual meeting of the American Association for Applied Linguistics, New York.

Diehl, W. (1980). *Functional literacy as a variable construct: An examination of the attitudes, behaviors, and strategies related to occupational literacy*. Unpublished doctoral dissertation, Indiana University.

Fisher, D.L. (1981). Functional literacy tests: A model of question-answering and an analysis of errors. *Reading Research Quarterly*, *16*, 418–448.

Guthrie, J.T. (1988). Locating information in documents: A computer simulation and cognitive model. *Reading Research Quarterly, 23*, 178–199.

Heath, S.B. (1980). The functions and uses of literacy. *Journal of Communication*, *30*, 123–133.

Jacob, E. (1982). *Literacy on the job: Final report of the ethnographic component of the industrial literacy project*. Washington, DC: Center for Applied Linguistics.

Kirsch, I.S., and Guthrie, J.T. (1984a). Adult reading practices for work and leisure. *Adult Education Quarterly, 34*(4), 213–232.

Kirsch, I.S., and Guthrie, J.T. (1984b). Prose comprehension and text search as a function of reading volume. *Reading Research Quarterly, 19*, 331–342.

Kirsch, I.S., and Jungeblut, A. (1986). *Literacy: Profiles of America's young adults—Final report* (NAEP Report No. 16–PL–Ol). Princeton, NJ: National Assessment of Educational Progress.

Kirsch, I.S., and Jungeblut, A. (1992). *Profiling the literacy proficiencies of JTPA and ES/UI populations: Final report to the Department of Labor*. Princeton, NJ: Educational Testing Service.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: US Department of Education.

Kirsch, I.S., Jungeblut, A., and Mosenthal, P.B. (1998). The measurement of adult literacy. In T.S. Murray, I.S. Kirsch, and L. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first international adult literacy survey*. Washington, DC: US Department of Education, National Center for Education Statistics.

Kirsch, I.S., and Mosenthal, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*, 5–30.

Kirsch, I.S., and Mosenthal, P.B. (1994). Interpreting the IEA Reading Literacy Scales. In M. Binkley, K. Rust, and M. Winglee (Eds.), *Methodological issues in comparative educational studies: The case of the IEA Reading Literacy Study*. Washington, DC: National Center for Education Statistics, United States Department of Education.

Lerner, D., and Lasswell, H.D. (1951). *The policy sciences: Recent developments in scope and method*. Stanford, CA: Stanford University Press.

Messick, S. (1987). Large-scale educational assessment as policy research: Aspirations and limitations. *European Journal of Psychology and Education, 2*(2), 157–165.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.

Messick, S., Beaton, A., and Lord, F. (1983). *National assessment of educational progress reconsidered: A new design for a new era*. (NAEP Report 83–1). Princeton, NJ: National Assessment of Educational Progress.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. *Reading Research Quarterly*, *17*(3), 400–419.

Miller, P. (1982). Reading demands in a high-technology industry. *Journal of Reading*, *26*(2), 109–115.

Montigny, G., Kelly, K., and Jones S., (1991). *Adult Literacy in Canada: Results of a National Study*. (Statistics Canada, Catalogue No. 89–525–XPE). Ottawa, Canada: Minister of Industry, Science and Technology.

Mosenthal, P.B., and Kirsch, I.S. (1989–1991). Understanding documents [monthly column]. *Journal of Reading*. Newark, DE: International Reading Association.

Mosenthal, P.B., and Kirsch, I.S. (1991). Toward an explanatory model of document process. *Discourse Processes*, *14*, 147–180.

Mosenthal, P.B., and Kirsch, I.S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent and Adult Literacy*, *41*(8), 638–657.

Organization for Economic Co-operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: Author.

Organization for Economic Co-operation and Development and Human Resources Development Canada. (1997). *Literacy skills for the knowledge society: Further results of the international adult literacy survey*. Paris, France, and Ottawa, Canada: Author.

Organization for Economic Co-operation and Development. (1992). *Adult illiteracy and economic performance.* Paris, France: Author.

Organization for Economic Co-operation and Development and Statistics Canada. (1995). *Literacy, economy and society: results of the first international adult literacy survey*. Paris, France, and Ottawa, Canada: Author.

Organization for Economic Co-operation and Development and Statistics Canada. (2000). *Literacy in the information age: Final report of the international adult literacy survey*. Paris, France, and Ottawa, Canada: Author.

Resnick, D., and Resnick, L. (1977). The nature of literacy—An historical exploration. *Harvard Educational Review*, *43*, 370–385.

Scribner, S., and Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Sticht, T.G. (Ed.). (1975). *Reading for working: A functional literacy anthology.* Alexandria, VA: Human Resources Research Organization.

Sticht, T.G. (1978). *Literacy and vocational competency* (Occasional Paper 39, National Center for Research in Vocational Education). Columbus, OH: Ohio State University.

Sticht, T.G. (1982, January). *Evaluation of the reading potential concept for marginally literate adults*. (Final Report FR–ET50–82–2). Alexandria, VA: Human Resources Research Organization.

Szwed, J. (1981). The ethnography of literacy. In M. Whitman (Ed.), *Writing: The nature, development, and teaching of written communication*: Vol. 1. Hillsdale, NJ: Erlbaum.

Tucker, L.R., and Finkbeiner, C.T. (1981). *Transformation of factors by artificial personal probability functions* (ETS Research Report 81–58). Princeton, NJ: Educational Testing Service.

Venezky, R.L. (1983). The origins of the present-day chasm between adult literacy needs and school literacy instruction. *Visible Language*, *16*, 113–136.

Werlich, E. (1976). *A text grammar of English*. Heidelberg: Quelle and Meyer.

Wickert, R. (1989). *No single measure*. Canberra, Australia: The Commonwealth Department of Employment, Education and Training.

Notes

1. This section is based on the work of Werlich, 1976. Category names in parentheses are alternative ways of labeling the class.
2. Mosenthal and Kirsch wrote a monthly column on Understanding Documents, which appeared in the *Journal of Reading* between 1989 and 1991.
3. While most of the tasks in IALS received common RP80 values, a few tasks were assigned values unique to a particular country when warranted by the data. Since the value assigned to each variable used in the regression analyses was based on the evaluation of each task in English, it was decided to use the RP80 values for the US as well.
4. The data used in the regression of prose literacy items are provided in Appendix B.
5. The data used in the regression of document literacy items are provided in Appendix B.
6. The data used in the regression of document literacy items are provided in Appendix B.

Appendix A

Coding rules for the process variables

Type of information

Type of information requested refers to the nature of information that readers must identify to complete a question or directive. Types of information form a continuum of concreteness, which was operationalized as follows for purposes of this analysis:

- When the requested information is a person, animal, place, or thing, score 1.
- When the requested information is an amount(s), time(s), attribute(s), action(s), or location(s), score 2.
- When the requested information is a manner, goal, purpose, condition, or predicate adjective, score 3.
- When the requested information is a cause, result, reason, evidence, similarity, or pattern, score 4.
- When the requested information is an equivalent, difference, or theme, score 5.

Plausibility of distracting information

Plausibility of distracting information refers to whether or not an identifiable match exists between information in the question and the text, or between the text and the distractors in a multiple-choice question, which makes it difficult for readers to identify the correct answer. The scoring rules for plausibility of distracting information are as follows:

- When there is no distracting information in the text, score 1.
- When distractors contain information that corresponds literally or synonymous to information in the text but not in the same paragraph as the answer, score 2.
- When distractors contain information that represent plausible invited inferences not based on information related to the paragraph in which the answer occurs, score 3.
- When one distractor in the choices contains information that is related to the information in the same paragraph as the answer, score 4.
- When two or more distractors in the choices contain information that is related to the information in the same paragraph as the answer, score 5.
- When one or more distractors represent plausible inferences based on information outside the text, score 5.

Type of calculation

The scoring rules for type of calculation are as follows:

- Score 1 if task requires a single addition.

- Score 2 if task requires a single subtraction.
- Score 3 if task requires a single multiplication.
- Score 4 if task requires a single division.
- Score 5 if task requires multiple operations.

Type of match

This variable relates to the nature of the task and the level of processing required to respond correctly to a task. The first diagram represents the additive scoring model used to code prose literacy tasks. It is followed by the model used to code document literacy tasks. The third model is for coding the variable "operation specificity" on the quantitative scale.

An additive scoring model for prose literacy tasks

If locate,	add 1;
If cycle,	add 2;
If integrate,	add 3;
If generate,	add 5;

If within paragraph,	add 0;
If between paragraphs,	add 1.

If infer condition is based on synthesis of features identified throughout paragraph, or if compare,	add 0;
If infer condition is based on synthesis of features identified between paragraphs, or if contrast,	add 1.

If 1 phrase to search on,	add 0;
If 2 phrases to search on,	add 1;
If 3 phrases to search on,	add 2,
If 4 phrases to search on,	add 3.

If 1 item response,	add 0;
If 2 item response,	add 1;
If 3-4 item response,	add 2;
If 5 or more item response,	add 3.

For multiple responses:

If number of responses is specified,	add 0;
If number of responses is unspecified,	add 1.

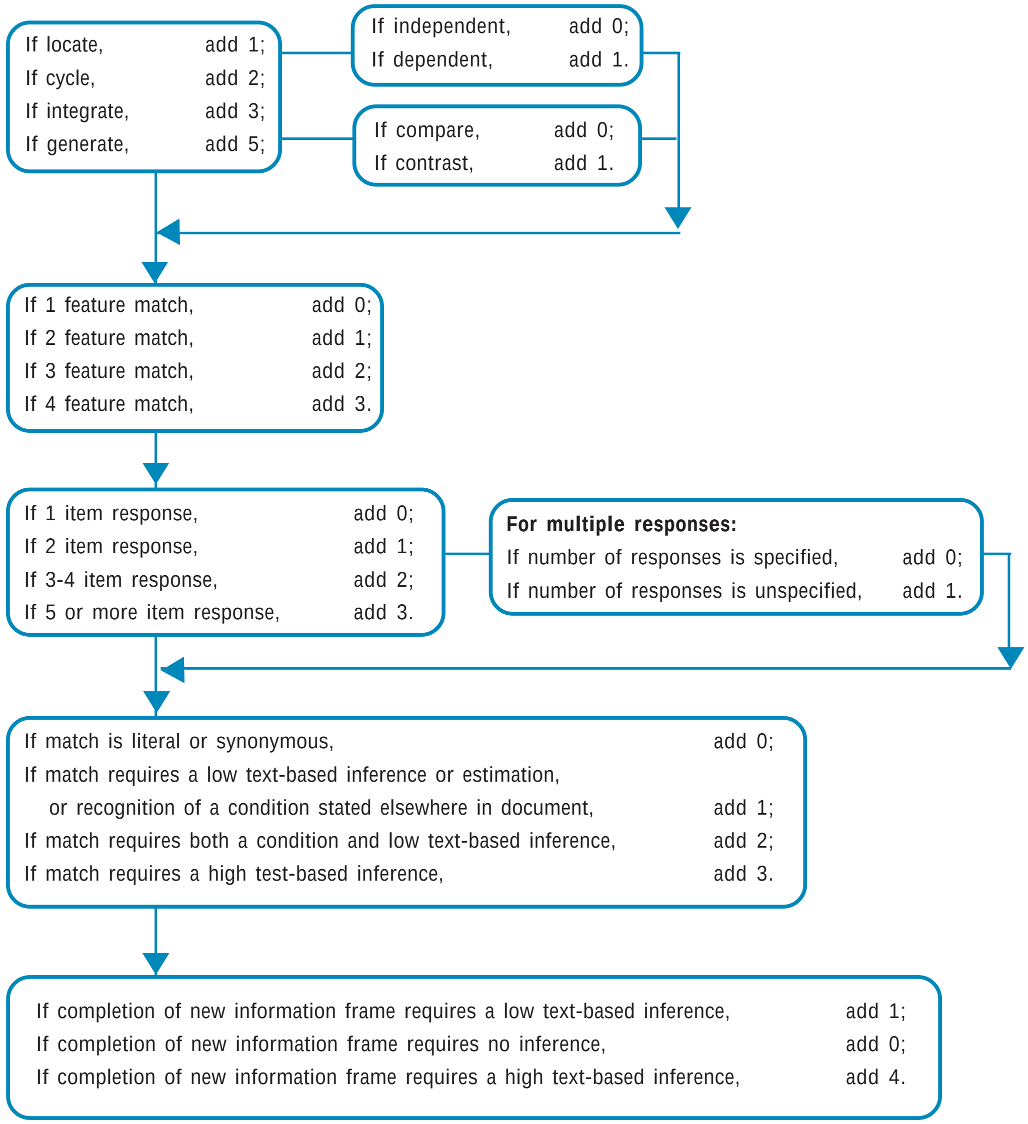
For given information:

If match is literal or synonymous,	add 0;
If match requires a low-level text-based inference,	add 1;
If match requires a high-level text-based inference,	add 3.

For requested information:

If completion of new information frame requires no inference or the identification of a paradigmatic relationship,	add 0;
If completion of new information frame requires a low-level text-based inference, identification of a condition or an antecedent, or restatement of type of information,	add 2;
If completion of new information frame requires some specialized prior knowledge or the identification of a syntagmatic relationship,	add 3;
If completion of new information frame requires a high-level text-based inference,	add 4.

An additive scoring model for document literacy tasks



Operation specificity

This variable deals with the extent to which the numbers are embedded in the text or document and the degree to which an inference must be made to identify the type of operation to be performed.

An additive scoring model for quantitative literacy tasks

If numbers are in a row or column format,	add 0;
If numbers are not in a row or column format,	add 2

If numbers are adjacent,	add 0;
If numbers are not adjacent,	add 1.

If labels or amounts are identified without a search,	add 0;
If labels are present and amounts are identified with a search,	add 1;
If labels are inferred and amounts are identified with a search,	add 2;
If one or more labels are ambiguous based on referents in question,	add 4.

If operation is signaled by +, -, X, /, or states 'add,' 'subtract,' 'multiply,' 'divide,' or 'total' (when it means add),	add 0;
If semantic relationship is stated, e.g. 'how much more,' 'how much less,' 'how many times,' 'calculate the difference'	add 1;
If operation is easily inferred; e.g., 'how much saved,' or 'deduct'	add 2;
If operation is based on known ratios; e.g., 'percent 0'	add 3.

If numbers are present,	add 0;
If numbers are entered or identified in previous task,	add 1;
If numbers are present but one is conditional,	add 2;
If numbers must be inferred but label is identified,	add 2.

If units require no transformation,	add 0;
If units require transformation (e.g., time or fraction), or require converting to common units, like decimals or fractions,	add 1.

Appendix B

Data from prose, document, and quantitative items

Table B1

Code values for IALS prose literacy items

Item	RP80	TOM	TOI	POD	READ
core1	190.00	1.00	2.00	1.00	6.00
b1q5	318.00	5.00	4.00	3.00	7.00
b1q6	297.00	3.00	3.00	3.00	7.00
b1q10	248.00	1.00	1.00	2.00	9.00
b1q11	377.00	7.00	3.00	4.00	9.00
b2q1	254.00	1.00	3.00	3.00	8.00
b2q3	230.00	1.00	4.00	2.00	8.00
b2q6	329.00	5.00	4.00	4.00	8.00
b2q7	374.00	6.00	4.00	4.00	8.00
b3q7	306.00	5.00	2.00	3.00	7.00
b3q8	338.00	6.00	5.00	1.00	7.00
b3q9	287.00	3.00	3.00	2.00	7.00
b3q11	281.00	3.00	3.00	2.00	8.00
b3q12	318.00	4.00	4.00	4.00	8.00
b3q13	297.00	3.00	3.00	2.00	17.00
b3q15	246.00	2.00	2.00	2.00	8.00
b4q1	188.00	1.00	2.00	1.00	6.00
b4q2	298.00	4.00	3.00	2.00	6.00
b4q6	314.00	3.00	3.00	2.00	8.00
b4q7	306.00	4.00	5.00	2.00	8.00
b5q1	192.00	1.00	1.00	1.00	8.00
b5q2	226.00	2.00	2.00	4.00	8.00
b5q3	255.00	1.00	2.00	3.00	12.00
b5q4	350.00	6.00	5.00	3.00	12.00
b5q5	324.00	3.00	3.00	4.00	12.00
b5q6	316.00	5.00	3.00	2.00	12.00
b6q1	209.00	1.00	4.00	1.00	6.00
b6q7	275.00	4.00	2.00	2.00	13.00
b6q8	310.00	5.00	3.00	3.00	13.00
b7q10	271.00	4.00	1.00	1.00	8.00
b7q11	349.00	7.00	3.00	2.00	8.00
b7q13	206.00	1.00	2.00	2.00	6.00
b7q14	294.00	5.00	2.00	3.00	6.00
b7q15	275.00	3.00	2.00	1.00	6.00

Table B2
Code values for IALS document literacy items

Item	RP80	TOM	TOI	POD	READ
core2	182.00	1.00	1.00	2.00	1.00
b1q1	291.00	2.00	3.00	2.00	2.00
b1q2	254.00	1.00	2.00	2.00	2.00
b1q13	237.00	1.00	2.00	3.00	2.00
b2q8	322.00	4.00	2.00	3.00	5.00
b2q10	304.00	3.00	2.00	4.00	4.00
b2q11a	231.00	1.00	2.00	1.00	4.00
b2q11b	280.00	4.00	2.00	3.00	4.00
b2q11c	227.00	1.00	1.00	1.00	4.00
b2q11d	221.00	1.00	1.00	1.00	4.00
b2q11e	237.00	2.00	1.00	1.00	4.00
b3q2	341.00	6.00	5.00	2.00	5.00
b3q5	296.00	4.00	2.00	3.00	7.00
b4q4	321.00	4.00	3.00	4.00	1.00
b4q5a	294.00	3.00	2.00	3.00	1.00
b4q12a	229.00	1.00	1.00	1.00	2.00
b4q12b	256.00	1.00	1.00	2.00	2.00
b4q12c	222.00	1.00	1.00	1.00	2.00
b4q12d	195.00	1.00	1.00	1.00	2.00
b5q7	242.00	2.00	2.00	2.00	5.00
b5q8	291.00	3.00	2.00	2.00	5.00
b5q10	295.00	5.00	4.00	3.00	5.00
b5q11d	302.00	4.00	2.00	2.00	3.00
b5q12	313.00	5.00	1.00	4.00	3.00
b6q4	218.00	1.00	1.00	2.00	2.00
b6q6	250.00	2.00	4.00	1.00	2.00
b6q9	270.00	3.00	2.00	3.00	9.00
b6q11	297.00	4.00	2.00	3.00	9.00
b7q1	188.00	1.00	2.00	2.00	2.00
b7q3	234.00	3.00	1.00	2.00	2.00
b7q4	270.00	4.00	2.00	4.00	6.00
b7q7	327.00	4.00	1.00	3.00	11.00
b7q8	287.00	4.00	2.00	2.00	11.00
b7q9	408.00	5.00	2.00	5.00	11.00

Table B3

Code values for IALS quantitative literacy items

Item	RP80-	TOM	POD	TOC	OSPEC	READ
Core	262.00	2.00	2.00	2.00	3.00	1.00
Core	232.00	1.00	1.00	1.00	1.00	1.00
b1q4	289.00	1.00	3.00	1.00	3.00	7.00
b1q7	300.00	2.00	2.00	2.00	4.00	6.00
b1q9	302.00	3.00	4.00	1.00	3.00	6.00
b1q14	327.00	3.00	4.00	5.00	4.00	2.00
b1q15	265.00	3.00	2.00	1.00	3.00	2.00
b2q4	315.00	2.00	3.00	5.00	5.00	2.00
b2q5	408.00	2.00	4.00	5.00	7.00	2.00
b2q9	255.00	4.00	4.00	2.00	3.00	5.00
b3q1	276.00	2.00	2.00	2.00	4.00	5.00
b3q3	277.00	3.00	4.00	1.00	2.00	5.00
b3q6	308.00	3.00	4.00	2.00	4.00	7.00
b3q14	328.00	1.00	1.00	3.00	6.00	17.00
b4q3	272.00	1.00	1.00	3.00	3.00	6.00
b4q5b	302.00	1.00	3.00	3.00	2.00	1.00
b4q9	324.00	3.00	4.00	2.00	3.00	11.00
b4q10	381.00	4.00	4.00	5.00	9.00	11.00
b4q11	280.00	4.00	3.00	1.00	4.00	11.00
b4q12	229.00	1.00	1.00	1.00	1.00	2.00
b4q12	225.00	1.00	1.00	1.00	1.00	2.00
b5q9	293.00	4.00	2.00	2.00	5.00	5.00
b5q11	336.00	1.00	1.00	1.00	6.00	3.00
b5q11	331.00	7.00	3.00	5.00	6.00	3.00
b5q13	335.00	5.00	5.00	1.00	4.00	5.00
b5q14	308.00	4.00	2.00	2.00	2.00	5.00
b6q2	315.00	3.00	3.00	5.00	3.00	6.00
b6q3	253.00	3.00	3.00	3.00	3.00	6.00
b6q5	287.00	2.00	2.00	1.00	3.00	2.00
b6q10	348.00	2.00	3.00	1.00	6.00	9.00
b7q2	268.00	2.00	2.00	2.00	5.00	2.00
b7q5	317.00	5.00	4.00	2.00	4.00	6.00
b7q6	321.00	1.00	3.00	3.00	3.00	6.00

Chapter 5

Adult Numeracy and its Assessment in the ALL Survey: A Conceptual Framework and Pilot Results

Iddo Gal	*University of Haifa, Israel*
Mieke van Groenestijn	*Utrecht University of Professional Education, Netherlands*
Myrna Manly	*El Camino College, California, USA*
Mary Jane Schmitt	*TERC, Cambridge, Massachusetts, USA*
Dave Tout	*Language Australia, Melbourne, Australia*

Acknowledgments

We thank the numerous international experts, study managers, and translators who have commented on earlier drafts of this framework or on sample items, or who have made suggestions.

Table of Contents

Table of Contents

Overview

About the ALL survey

The Adult Literacy and Lifeskills (ALL) survey (formerly the International Life Skills Survey, ILSS), is a follow-up to the International Adult Literacy Survey (IALS), the world's first large-scale comparative assessment of adult literacy. The ALL survey has been a joint development by Statistics Canada and by the United States' National Center for Education Statistics (NCES), in cooperation with the Organization for Economic Cooperation and Development (OECD) and the United Nations Educational, Scientific, and Cultural Organization (UNESCO (OREALC)) since 1997.

Using household survey methods, the ALL project assesses performance of adults aged 16 to 65 in the domains of *Prose* and *Document Literacy*, *Numeracy*, and *Problem Solving*, and also collects information about experiences with Information and Communication Technology. Many other variables and correlates of interest, such as demographic details, employment status, and participation in learning activities, are recorded via a Background Questionnaire (BQ).

The key goals of the project are to:

- Profile and compare skill distributions across participating countries
- Explore covariates of observed skills, e.g., social and economic factors,
- Understand the relationship of Numeracy and Problem Solving to Prose and Document Literacy.

Following several years of preparation and an extensive Pilot survey in 2002, the first round of ALL's main assessment, involving 6 countries, began in 2003. Plans include a full comparative report from the first round, to be followed by other publications and by assessment rounds in additional countries.

Numeracy in the ALL survey

Numeracy is included as a domain in the ALL Survey as one of the critical factors in determining the capability of a population to adapt to and effectively function in an increasingly information-laden society or to perform well at work (European Commission, 1996). Schools are placing more emphasis on the links between the knowledge and skills gained in the mathematics classroom and students' ability to handle real-life situations that require activation of mathematical knowledge and skills. Given the increasing need for adults to continuously adapt to changing citizenship, workplace, and everyday life demands, it is vital that nations have information about their workers' and citizens' numeracy in order to evaluate the human capital available for advancement, to plan effective school-based and lifelong learning opportunities, and to better understand the factors that affect citizens' ability to advance their well-being.

The conception of numeracy developed for ALL is built upon recent research and work done in several countries on functional demands of different life contexts, on the nature of adults' mathematical and statistical knowledge and skills, and on how such skills are applied or used in different circumstances. In light of the general intention of the ALL survey to provide information about a diverse set of lifeskills, this framework defines numeracy as follows:

> *Numeracy is the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations.*

However, since an assessment can only examine observed behavior, not internal processes or capacities, this framework uses a more detailed definition of "numerate behavior" as a means to guide the development of items for the survey.

> *Numerate behavior is observed when people manage a situation or solve a problem in a real context; it involves responding to information about mathematical ideas that may be represented in a range of ways; it requires the activation of a range of enabling knowledge, factors, and processes.*

Both definitions above use broad concepts and terms whose specific meanings and underlying components are explained in more detail later in this document. The definitions imply that numeracy can be viewed as a functional competency that is somewhat different from the traditional notion of "knowing mathematics" in that it relates to the capacity to act and bring one's knowledge (mathematical and other) to bear on tasks in context.

Numeracy is assessed in the ALL survey by presenting to respondents short tasks with mathematical content that are embedded in real-life contexts. As illustrated by the sample items in Appendix 1, tasks require that respondents activate a range of knowledge and skills, and respond to different situations by computing, estimating, understanding notions of shape, length, volume, monetary units, measuring, understanding some statistical ideas, or interpreting simple formulas. Respondents are encouraged to use the tools provided, paper and pencil, a ruler and a four-function calculator, whenever they wish. Numeracy and its assessment is thus broader in scope than Quantitative Literacy, as defined in IALS, which refers to a person's ability to apply arithmetic operations to numbers embedded in print materials.

About this report

This report describes key stages in the development of the Numeracy assessment scale for the ALL survey that took place between 1998 and early 2003, in three parts:

Part A presents a conceptual discussion of the numeracy construct and its facets, examines approaches to assessment of mathematical skills, and reviews issues that influenced the item development process. This section includes most of the text that appeared in the original Numeracy framework published in 1999 on the ALL (then ILSS) website. Editorial changes were made to streamline the presentation and to respond to comments from external reviewers, but the material was kept mostly intact since it served as the conceptual foundation from which item development progressed starting in 1998.

Part B describes the development of the item pool and scoring guidelines, and the feasibility studies that led to the selection of 80 items for a Pilot study that took place in 2002.

Part C outlines the design of the ALL Pilot study and presents key results, on the basis of which 40 Numeracy items were selected for the main ALL assessment in 2003. Two *appendices* contain sample Numeracy items that have been released to the public (Appendix 1) and details of a scheme of complexity factors that was used to inform the evaluation of difficulty levels of items during the item development process (Appendix 2).

Part A:

Conceptual foundations

This part of the report presents a framework for the assessment of numeracy skills in the Adult Literacy and Lifeskills Survey (ALL). It begins by examining different perspectives that informed the conceptualization of numeracy developed for the ALL survey, and reviews some factors known to affect adults' numerate behavior, such as literacy, beliefs and attitudes, and prior practices, some of which are studied by specific questions in the ALL Background Questionnaire (BQ). Next, a definition of numeracy is presented, followed by an elaboration on the facets of numerate behavior that guided the development of items for the ALL Numeracy scale. Finally, several factors that are thought to influence the complexity or difficulty of numeracy tasks are presented; these factors are important both for development of items as well as for interpretation of the survey results. The actual stages of item development and pilot testing are described in parts B and C of the report.

1. Adult numeracy: Influences and perspectives

The construct "numeracy" does not have a universally accepted definition, nor agreement about how it differs from "mathematics." For some educators and officials, the term numeracy implies a set of simple skills involving the ability to carry out basic computations or arithmetical operations. In contrast, adult numeracy as viewed here is a broad construct that involves a range of knowledge, skills, and supporting processes. This section presents a review of influences or trends that have led to the conceptualization of numeracy employed in the ALL survey, and organizes them under five headings: *Workplace perspectives*, *Broader life purposes*, *Educational perspectives*, *Research perspectives*, and *Assessment schemes*. (These five headings or areas are interconnected and not mutually exclusive, as developments in one area often influence thinking in other areas).

1.1 Workplace perspectives

Over the last two decades, many countries have launched large-scale efforts aimed to define "core skills" or "key competencies" that workers should have, in response to the need to maintain economic competitiveness and improve employability of adults and school graduates. As workplaces are becoming more concerned with involving all workers in improving workplace efficiency and quality processes, the importance of numeracy skills is growing, and they have been shown to be a key factor in workplace success (Jones, 1995; Murnane, Willett and Levy, 1995). Basic computational knowledge has always been considered as part of the fundamental skills that adults need to possess, but the recent skills frameworks describe in specific terms the need for workers to possess a much broader range of mathematical skills. Examples exist in many countries, such as the United Kingdom, the United States, Australia, the Netherlands, and other OECD countries. The following descriptions are indicative of the nature of such efforts.

Outcomes from skills projects conducted in the United States illustrate workplace perspectives regarding mathematical skills needed by workers. Following earlier research by a task force of the American Society of Training and Development (Carnevale, Gainer, and Meltzer, 1990), the Secretary of Labor's Commission on Achieving Necessary Skills (SCANS) (Packer, 1997) has differentiated between mastery of basic arithmetical skills and much broader and flexible understanding of principles and underlying ideas, subsumed under the notion of mathematical skills:

> *SCANS arithmetical skills:* Performs basic computations; uses basic numerical concepts such as whole numbers and percentages in practical situations; makes reasonable estimates and arithmetic results without a calculator; and uses tables, graphs, diagrams and charts to obtain or convey quantitative information.
>
> *SCANS mathematical skills:* Approaches practical problems by choosing appropriately from a variety of mathematical techniques; uses quantitative data to construct logical explanations for real world situations; expresses mathematical ideas and concepts orally and in writing; and understands the role of chance in the occurrence and prediction of events. (SCANS, 1991, p. 83)

Based on a later survey of employers, industry trainers, and educators, among others, Forman and Steen (1999) similarly argued that quantitative skills desired by employers are much broader than mere facility with the mechanics of addition, subtraction, multiplication, and division and familiarity with basic number facts; they also include some knowledge of statistics, probability, mental computation strategies, some grasp of proportional reasoning or modeling relationships, and broad problem-solving and communication skills about quantitative issues. Buckingham (1997), who studied what she called "specific and generic numeracies of the workplace" in some manufacturing industries in Australia, concluded that workplace numeracy is now about making decisions in the face of uncertainty in real situations, and that it encompasses far more than the basic skills traditionally associated with the term numeracy (as this term had been used in Australia).

Outcomes from skills projects are echoed in educational specifications. For example, basic skills projects in the United Kingdom and Australia influenced vocational education frameworks that name numeracy as an important skill, and describe stages or levels of accomplishment (Australian Education Council, 1992). The National Council for Vocational Qualifications Core Skills in the United Kingdom identified five levels of numeracy skill (Oates, 1992) that are closely linked to the sequence of content in the national school mathematics curriculum.

1.2 Broader life purposes

Since numeracy involves action in the real world, it is important to reflect on the kinds of purposes served by people's ability to act in a numerate way. Since people's numeracy is related to and may at times depend on people's literacy skills or other lifeskills, the purposes served by numeracy are expected to parallel those served by adults' literacy.

Work to describe the purposes served by adults' literacy and numeracy skills has been conducted in several countries. In Australia, for example, Kindler et al., (1996) reported on four such purposes: literacy for self-expression, literacy for practical purposes, literacy for knowledge, and literacy for public debate. In the Equipped for the Future initiative, The National Institute for Literacy in the United States has sponsored efforts to define critical skill areas. As part of the project, adult learners were asked what they needed to compete in a global economy and exercise the rights and responsibilities of citizenship. Content analysis yielded four broad types of purposes (Stein, 1995):

- Literacy for access and orientation in the world,
- Literacy as voice to one's ideas and opinions,
- Literacy for independent action, solving problems and making decisions as a parent, citizen and worker,
- Literacy as a bridge to further learning and to keep up with a rapidly changing world.

In Australia, a range of work has been done to create standards and a hierarchy of numeracy skill development that is not based upon school mathematics descriptions (Coates et al. 1995). In one key project (Kindler et al. 1996), numeracy was organized into four broad categories, according to different purposes and functions of using mathematics. *Numeracy for Practical Purposes* addresses aspects of the physical world that involve designing, making, and measuring. *Numeracy for Interpreting Society* relates to interpreting and reflecting on numerical and graphical information in public documents and texts. *Numeracy for Personal Organization* focuses on the numeracy requirements for personal organizational matters involving money, time and travel. *Numeracy for Knowledge* describes the mathematical skills needed for further study in mathematics, or other subjects with mathematical underpinnings and/or assumptions.

Overall, the purposes regarding literacy and numeracy appear to agree and suggest that adults need to be able to apply their numeracy and literacy skills to tasks with a social purpose in both informal and more formal contexts.

1.3 Educational perspectives

Recent years saw a growing dialogue about the goals and impact of mathematics education in schools. Various arguments have been brought forward to support a broadening of the conceptions regarding the mathematical skills and knowledge that school graduates should possess. In a society in which the media constantly present information in numerical or graphical form, the ability to interpret quantitative and statistical messages has been positioned by key stakeholders in education as vital for all adults (Steen, 1997). While employers have focused mostly on practical or job-specific numeracy skills, educators associated with the mathematical sciences have also paid much attention to the importance of quantitative literacy in civic and social contexts, and argued that mathematics is a crucial part of a common fabric of communication indispensable for modern civilized society, in part because it is the language of science and technology. Thus, understanding of public discussions and reports about socially important topics such as health and environmental issues are impossible without using the language of mathematics (National Research Council, 1989).

More mathematics educators now encourage links between the knowledge and skills gained in the mathematics classroom and students' ability to handle real-life situations that require activation of mathematical knowledge and skills (National Council of Teachers of Mathematics (NCTM),1989; Willis, 1990; Heuvel-Panhuizen and Gravemeijer, 1991). Such "handling" should be interpreted broadly, to mean not only application of mathematical procedures and concepts, but also many other abilities, such as the ability to critically reflect on information encountered (Frankenstein, 1989) or to understand and appreciate mathematical phenomena in the world, such as symmetry in the arts and nature.

The Realistic Mathematics Education (RME) initiative in the Netherlands that started in the early 1980s is an example of an attempt to develop educational experiences in light of the characteristics of real-world mathematical tasks and practices. RME is based upon the assumption that mathematics is an essential and important aspect of society, and therefore that mathematics education should be derived from real-life situations and should aim to create those skills applicable in any societal situation (family, work, etc). RME in adult education aims to optimize mathematical knowledge, skills, and problem-solving strategies that people have already been using in everyday life, or learned in or out of school, so that they can apply those strategies flexibly in all kinds of situations, have more control over their own personal, societal and work lives, and undertake further learning or training. Therefore, RME implies that school students and adults are to also be trained in cooperative learning and in recognizing and facilitating their own problem-solving procedures, strategies, and learning processes (van Groenestijn, 1998).

In the adult education sector, which is growing and becoming more formalized in many countries due to economic considerations and the need for lifelong learning, attention to mathematical skills is very visible (Benn, 1997). Educators working with adults aim to assist learners in developing mathematical concepts and relationships in ways that are personally meaningful but also functional. Adult educators usually assume that there is rarely only one right way, but a wide variety of strategies that work well when solving functional computational problems. Adults' personal methods of using mathematics are encouraged and valued. This is often a significant difference from traditional (pre-reform) school-based mathematics teaching, within which school students were often expected to follow the one correct method, or algorithm, introduced by the teacher to solve a problem.

The National Institute for Literacy in the United States has sponsored several efforts to define critical skill areas, as part of its Equipped for the Future initiative. One key project, by the Adult Numeracy Network (ANN) (Curry, Schmitt, and Waldron, 1996), was designed to reach a consensus on the kinds of mathematics that adults should know and hence are important to teach and assess in adult education. This project aimed to consolidate several curricular perspectives, mainly those offered by the NCTM (1989) the SCANS Commission (1991), and prior work by the ABE Mathematics Team in Massachusetts (Leonelli, Merson, Schmitt, and Schwendeman, 1994), as well as the results of interviews with hundreds of adult learners, numeracy teachers, and employers.

The ANN's *Framework for Adult Numeracy Standards: The Mathematical Skills and Abilities Adults Need to be Equipped for the Future*, organized needed knowledge into seven broad themes or areas: Relevance/connections, Problem solving/reasoning/decision making, Communication, Number and number sense, Data, Geometry: spatial sense and measurement, and Algebra: patterns and functions. The first three themes are concerned with processes of being numerate, while the latter four cover key content areas of mathematics. Again, this framework highlights and supports the view that numeracy is about making meaning of mathematical information, and that it encompasses a broad spectrum of skills and knowledge bases.

1.4 Research perspectives

A sizeable literature has accumulated over the last several decades regarding the ways in which adults use mathematical skills or cope with mathematical tasks in both formal (i.e., school-based) and informal (i.e., everyday or workplace) contexts. However, few attempts have been made to synthesize this literature and examine its implications for large-scale skills assessments. Some examples of research reports and theoretical discussions that could be considered in this regard include: Rogoff and Lave, 1984; Resnick, 1987; Saxe, 1988; Carraher, Schliemann, and Carraher, 1988; Scribner and Sachs, 1991; Nunes, 1992; or Coben, 2000.

Based on an analysis of the above and related literature, Gal (1993), while at the Numeracy Project at the National Center for Adult Literacy in the US, developed a conceptual perspective on the nature of adults' numeracy and numerate behavior. Gal (1993; 1997) argued that "numeracy" refers to the aggregate of skills, knowledge, and dispositions that enable and support independent and effective management of diverse types of quantitative situations. Gal further argued that the scope of adult numeracy is broad in light of the need for different types of responses in different situations.

Some situations call for generative responses, i.e., computing a number or generating an estimate or a decision. Examples are dealing with simple operations (measuring the length of a shelf), dealing with multi-step operations embedded in text (such as completing a tax form) and making reasonable decisions (for example, choosing the best loan). Other situations call for interpretive responses, i.e., making sense of quantitative statements or data displays (as in a newspaper article reporting crime statistics), and being able to ask critical questions about the information and arguments presented without performing any calculations. Both types of situations, and many mixed types, vary in terms of the literacy and communication skills they involve; in some cases it may not be possible to separate literacy from numeracy skills. It has also been suggested (Gal, 1997) that numerate behavior is enabled by *dispositional elements* (beliefs, attitudes, habits of mind) that motivate and support effective behavior in any given situation.

Gal (1993) also proposed that adults *manage* situations that call for application of numeracy skills. A person may decide to sacrifice precision or accuracy to reduce mental load or save time. A response may be reached in a computationally inefficient way or be based on non-standard procedures, but this may not matter in real-life as long as the individual expends time and effort in a way that is reasonable in light of the demands of the situation and his or her goals. It follows that there may be important differences between how adults respond to a school-oriented task (where adults may try to apply only school-based, memorized procedures), and demonstrate numerate, confident behavior in realistic situations. Cumming, Gal, and Ginsburg (1998) have argued that many of these aspects of numerate behavior are not reflected in how tests and test items are created and interpreted.

1.5 Assessment schemes

Some understanding of the mathematical needs of adult life can also be gleaned from an examination of large-scale assessment efforts, used either with adults or school students.

Adult assessments. A framework developed by Kirsch and Mosenthal (see Kirsch, Jungblut, and Mosenthal, 1998) to describe adults' literacy skills, including aspects of adult's quantitative skills, has been widely implemented in multiple national and international assessment projects, most recently the International Adult Literacy Survey (IALS; see Statistics Canada and OECD, 1996, 1997). The IALS framework made use of three literacy scales—Prose Literacy, Document Literacy, and Quantitative Literacy—to operationalize its conception of literacy. The ALL domain of numeracy is most closely related to the Document Literacy (DL) and Quantitative Literacy (QL) scales, which were defined as follows.

DL: *The knowledge and skills required to locate and use information contained in various formats (including job applications, payroll forms, transportation schedules, maps, tables, and graphics).*

QL: *The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials (such as balancing a check book, figuring out a tip, completing an order form, or determining the amount of interest on a loan).*

QL tasks as well as some DL tasks have addressed important aspects of people's mathematical knowledge and skills. For example, DL tasks required respondents to identify, understand, and interpret information given in various lists, tables, charts and displays; this information sometimes included quantitative information, such as numbers or percents. QL tasks required respondents to apply arithmetical operations learned mostly in elementary grades. However, these tasks did not require respondents to cope with other types of mathematical information (e.g., measurements, shapes) or with information whose processing does not require comprehension of text. In addition, tasks used in both scales called for a limited range of responses, i.e., exact computations or specific types of interpretations. Thus, while such tasks and responses are important by themselves, they represent only a subset of the much wider range of tasks and responses that are typical of many everyday and work tasks, such as sorting, measuring, estimating, conjecturing, or using models (e.g., formulas).

School assessments. IALS' central mission was assessing facets of real-world literacy, hence the QL scale focused on application of basic mathematical operations in response to functional tasks using realistic texts. Large-scale assessments of mathematical skills aimed at younger populations usually take quite different approaches. Selected such assessments, associated with the ongoing PISA (Project for International Student Assessment), organized by the OECD, and with the GED test in the US, are reviewed below. These assessments are reviewed to highlight issues that can inform the content of an assessment of adult numeracy skills, as well as to shed light on areas where an assessment of adult numeracy has to deviate from familiar forms of assessments that are common with school-age populations.

The PISA survey has developed a framework for assessing Mathematical Literacy, defined as follows:

An individual's ability to identify, to understand, to make well-founded judgments about, and to act towards the roles that mathematics plays in dealing with the world, as needed for that individual's current and future life as a constructive, concerned, and reflective citizen (PISA Mathematics Functional Expert Group, 1998).

This definition shows some overlap and consistency with the conception of numeracy used in the present framework as well as with broader conceptions of literacy as adopted by IALS and ALL. Yet, some key differences seem to include the following:

- PISA focuses on how students understand, use, and apply mathematical skills and mathematize problems that are related to the *formal* school mathematics curriculum the students were expected to cover as part of their studies.

- PISA puts only partial emphasis on the realism of tasks. Given that students have limited world experience, tasks can be contrived or use formal symbolism that assesses mostly formal knowledge of what was taught in schools.

- The PISA mathematical assessment is not explicitly interested either in tasks where mathematical information is embedded in text (realistic or otherwise), or in the influence of literacy skills on mathematical performance (despite the inclusion of the term "literacy" in "mathematical Literacy").

Nationally-recognized standardized tests used in several countries to assess the mathematical knowledge of *adults* are often in line with school-based assessments. For example, the GED test in the U.S. (used to grant a high-school equivalency diploma to adults who did not formally graduate from high-school), and the National Vocational Qualifications system in the U.K., both use items with characteristics that are more in line with school-related assessments of mathematical knowledge than with the QL scale. These tests rely heavily on multiple-choice questions, employ some tasks requiring manipulation of numbers without a meaningful context, and require the use of some formal mathematical notations in formulas, either memorized or provided as part of the test.

School-oriented assessments point to some general areas of mathematical knowledge and skill that both school graduates as well as early school leavers may need to have to effectively cope with the various challenges of adult life. Reviewing the PISA, the *Third International Mathematics and Science Study* (TIMSS), and similar assessments highlights the fact that some important mathematical skills and knowledge that these assessments aim to capture were not captured by the QL scale of IALS. For example, knowledge of "big ideas" related to shape and geometry or to chance and statistics, knowledge of measurement systems, or the ability to "model" the mathematical aspects of certain situations were not included.

The above discussion is not meant to be a comprehensive review of current large-scale assessments of schooling-related mathematical skills (see Robitaille and Travers, 1992), nor a criticism of the QL scale of the IALS assessment framework. It simply reiterates that *all* assessments make conscious decisions regarding the (mathematical) skills that are important to assess, and that consequently the *forms* of assessment chosen carry not only advantages, but also disadvantages. The philosophy behind the design of mathematical assessments for PISA, GED, and similar assessments is based on assumptions about what it means to "know math" or "be able to do math" in a *schooling* context; hence, the assessment design assumes that it is legitimate to use a certain degree of formalization of math symbols and to present contrived math tasks. This assumption does not fit the assessment of skills of adults who may have been out of school for many years.

2. Towards a definition of Numeracy for ALL

The discussion above implies that adult numeracy should be viewed as different from "knowing school mathematics", that it is broader than the construct of Quantitative Literacy as defined in IALS, and also that multiple factors affect the way adults cope with the demands imposed by tasks that contain mathematical elements. (Note: For convenience, this report uses the term "mathematical" as inclusive of situations where *statistical* or *probabilistic* information may appear or where statistical thinking is required as well, even though statistical reasoning is not usually viewed as a branch of mathematics.).

Although a universally accepted definition of "numeracy" does not exist (Baker and Street, 1994), it is instructive to further examine some definitions and perspectives on the meaning of numeracy, and note that all contain an emphasis on the practical or functional application and use of mathematical knowledge and skills. The Australian Beazley Committee definition is typical:

> *Numeracy is the mathematics for effective functioning in one's group and community, and the capacity to use these skills to further one's own development and of one's community (Beazley, 1984).*

Another important element in defining numeracy is that of the role of communication processes. Numeracy not only incorporates the individual's abilities to use and apply mathematical skills efficiently and critically, but also requires the person to be able to interpret textual or symbolic messages as well as communicate about mathematical information and reasoning processes (Marr and Tout, 1997; Gal, 1997).

Most recent definitions of numeracy explicitly state that numeracy does not only refer to operating with numbers, as the word can suggest, but covers a wide range of mathematical skills and understandings. In recent years there has been much discussion and debate about the relationship between mathematics and numeracy and about the concept of "critical" numeracy (Frankenstein, 1989, Johnston, 1994). Johnston, for example, has argued that:

> *To be numerate is more than being able to manipulate numbers, or even being able to 'succeed' in school or university mathematics. Numeracy is a critical awareness which builds bridges between mathematics and the real-world, with all its diversity (Johnston, 1994).*

Many authors also argue that a discussion of functional skills should also address supporting or enabling attitudes and beliefs. In the area of adults' mathematical skills, "at homeness" with numbers or "confidence" with mathematical skills is expected, as these affect how skills and knowledge are actually put into practice (Cockroft, 1982; Tobias, 1993).

A definition of numeracy that seems to incorporate several of the aspects of numeracy noted above is from the Queensland Department of Education (1994) in Australia:

> *Numeracy involves abilities that include interpreting, applying and communicating mathematical information in commonly encountered situations to enable full, critical and effective participation in a wide range of life roles.*

An important commonality in the above descriptions of numeracy is the presence of mathematical elements in real situations, and the notion that these can be used or addressed by a person in a goal-oriented way, dependent on the needs and interests of the individual within the given context (home, community, workplace, etc.), as well as on his or her dispositions. Our earlier discussions further imply that numeracy involves more than just arithmetical skills as assessed in IALS, but extend to possession of number sense, estimation skills, measurement, and to multiple ways of responding flexibly to a mathematical situation. Finally, given the extent to which numeracy pervades the modern world, it is not necessarily just commonly encountered situations that require numerate behavior, but also *new* situations.

With the above in mind, a brief definition of numeracy proposed for the ALL is:

> *Numeracy: The knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations.*

This brief definition of numeracy, and the fuller definition of "numerate behavior" provided in the next subsection, are much broader than the definition of QL as used in the IALS (see section 1.5). Its key concepts relate in a broad way to situation management and to a range of effective responses (not only to application of arithmetical skills). It refers to a wide range of skills and knowledge (not only to computational operations) and to a wide range of situations that present actors with mathematical information of different types (not only those involving *numbers* embedded in *printed* materials).

The brief definition above implies that numerate individuals are those who respond autonomously to situations in which mathematical ideas are embedded, actively using the power of mathematics rather than delegating or ignoring quantitative issues. However, a broad, inclusive definition alone is not enough to fully describe what a numerate person can do. The next section elaborates on specific facets of numerate behavior that underlie the general terms used in the above definition. These facets provide the detail necessary to develop an assessment tool to evaluate numeracy in adult populations.

3. Facets of numerate behavior

To develop a guide for the production of items, the main challenge was to determine how to assess, within the constraints of the ALL Survey protocol, the extent to which an adult effectively manages and responds to the mathematical demands of diverse situations. The numeracy team had to find ways to bound the range of tasks for assessing "numeracy" in a way that captures the breadth of the concept and yet is pragmatic for a large-scale assessment. Focusing on the idea of measurable numerate behavior was key to solving the challenge.

In actuality, people's numeracy is revealed through the responses or behaviors they generate (i.e., identifying, interpreting, acting upon, communicating) in reaction to the mathematical information or ideas that may be represented in a situation. As implied by the literature and ideas reviewed earlier in Section 1, the nature of a person's responses to the mathematical and other demands of a situation will depend critically on the activation of various enabling knowledge bases (understanding of the context; knowledge and skills in the areas of mathematics, statistics, and literacy), on reasoning processes, and on certain dispositions. It is clear that numerate behavior will involve an attempt to engage with a task and not delegate it to others or deal with it by intentionally ignoring its mathematical content.

Thus, a definition of numerate behavior that underlies the assessment of Numeracy in the ALL Survey is:

> *Numerate behavior is observed when people manage a situation or solve a problem in a real context; it involves responding to information about mathematical ideas that may be represented in a range of ways; it requires the activation of a range of enabling knowledge, factors, and processes.*

Table 1 lists specific components of five key facets of numerate behavior. These facets and their components are further explained in subsequent sections.

3.1 Facet 1: Contexts

People try to manage or respond to a numeracy situation because they want to satisfy a purpose or reach a goal. Four types of purposes and goals are described below. To be sure, these are not mutually exclusive and may involve the same underlying mathematical themes.

Table 1:

Numerate behavior and its five facets

Numerate behavior involves:

1. managing a situation or solving a problem in a real context:
   everyday life
   work
   societal
   further learning
2. by responding:
   identifying or locating
   acting upon:
   - order/sort
   - count
   - estimate
   - compute
   - measure
   - model

   interpreting
   communicating about
3. to information about mathematical ideas:
   quantity and number
   dimension and shape
   pattern and relationships
   data and chance
   change
4. that is represented in a range of ways:
   objects and pictures
   numbers and symbols
   formulae
   diagrams and maps
   graphs
   tables
   texts
5. and requires activation of a range of enabling knowledge, factors, and processes:
   mathematical knowledge and understanding
   mathematical problem-solving skills
   literacy skills
   beliefs and attitudes.

Everyday life. The numeracy tasks that occur in everyday situations are often management tasks that one faces in personal and family life. Others revolve around hobbies, personal development, and interests. Representative tasks are handling money and budgets, comparison shopping, personal time management, making decisions involving travel, planning holidays, mathematics involved in hobbies like quilting or wood-working, playing games of chance, understanding sports scoring and statistics, reading maps, and using measurements in home situations such as cooking or home repairs.

Work-related. At work, one is confronted with quantitative situations that often are more specialized than those seen in everyday life. In this context, people may develop good skills in managing situations that might be narrower in their application of mathematical themes. Representative tasks are completing purchase orders, totaling receipts, calculating change, managing schedules, budgets, and project resources, using spreadsheets, organizing and packing different shaped goods, completing and interpreting control charts, making and recording measurements, reading blueprints, tracking expenditures, predicting costs, and applying formulas.

Societal or community. Adults need to know about trends and processes happening in the world around them (e.g., regarding crime, health issues, wages, pollution) and may have to take part in social events or community action. This requires that adults can read and interpret quantitative information presented in the media, including statistical messages and graphs. Also, they may have to manage situations like organizing a fund-raiser, realizing the fiscal effect of community programs, or interpreting the results of a study of the latest health fad.

Further learning. It is often also important to have numeracy skills that enable a person to participate in further study, whether for academic purposes or as part of vocational training. In either case, it is important to be able to know some of the more formal aspects of mathematics that involve symbols, rules, and formulas and to understand some of the conventions used to apply mathematical rules and principles.

3.2 Facet 2: Responses

In different types of real-life situations, people may have to respond in one or more of the following ways (the first virtually always occurs; others will depend on the interaction between situational demands and the goals, skills, dispositions, and prior learning of the person):

Identify or locate some mathematical information present in the task or situation confronting them that is relevant to their purpose or goal.

Act upon or react to the information in the situation. Bishop (1988), for example, proposed that there are six modes of mathematical actions that are common in all cultures: counting, locating, measuring, designing, playing and explaining. Other types of actions or reactions may occur, such as doing some calculations ("in the head" or with a calculator), ordering or sorting, estimating or modeling (such as by using or developing a formula).

Interpret the information embedded within the situation (and the results of any prior action) and comprehend what it means or implies. This can include making a judgment about how mathematical information or known facts actually apply to the situation or context. Contextual judgment may have to be used in deciding whether an answer makes sense or not in the given context, for example, that a result of "2.35 cars" is not a valid solution to how many cars are needed to transport a group. It can also incorporate a critical aspect, where a person questions the purpose of the task, the validity of the data or information presented, and the meaning and implications of the results, both for them as an individual and possibly for the wider community.

Communicate about the mathematical information given, or the results of one's actions or interpretations to someone else. This can be done orally or in writing (ranging from a simple number or word to a detailed explanation or analysis) and/or through drawing (a diagram, map, graph).

3.3 Facet 3: Mathematical information

Mathematical information can be classified in a number of ways and on different levels of abstraction. One approach is to refer to fundamental "big ideas" in the mathematical world. Steen (1990), for example, identified six broad categories pertaining to: *Quantity*, *Dimension*, *Pattern*, *Shape*, *Uncertainty*, and *Change*. Rutherford & Ahlgren (1990) described networks of related ideas: *Numbers*, *Shapes*, *Uncertainty*, *Summarizing data*, *Sampling*, and *Reasoning*. Dossey (1997) categorized the mathematical behaviors of quantitative literacy as: *Data representation and interpretation*, *Number and operation sense*, *Measurement*, *Variables and relations*, *Geometric shapes and spatial visualization*, and *Chance*. The ALL Numeracy team drew from these three closely tied categorizations to arrive at a set of five fundamental ideas that in their view characterize the mathematical demands met by adults in diverse situations at the beginning of the 21st century.

Quantity and Number. *Quantity* is described by Fey (1990) as an outgrowth of people's need to quantify the world around us, using attributes such as: length, area, and volume of rivers or land masses; temperature, humidity, and pressure of our atmosphere; populations and growth rates of species; motions of tides; revenues or profits of companies, etc. *Number* is fundamental to quantification and different types of number constrain quantification in various ways: whole numbers can serve as counters or estimators; fractions, decimals and percents as expressions of greater precision, parts or comparisons (ratios); and positive and negative numbers as directional indicators. In addition to quantification, numbers are used to put things in order and as identifiers (e.g., telephone numbers or zip codes). Facility with quantity, number, and operation on number requires a good "sense" of magnitude. Contextual judgment comes into play when deciding how precise one should be or which tool (calculator, mental math, a computer) to use. Money and time management, the ubiquitous mathematics that is part of every adult's life, depends on a good sense of number and quantity. A basic level numeracy task might be figuring out the cost of one can of soup, given the cost of 4 for $2.00; a task with a higher cognitive demand could involve more complex numbers such as when figuring out the cost per pound when buying 0.783 kg of cheese for 12,95 Euros.

Dimension and shape. *Dimension* includes "big ideas" related to one, two, and three dimensions of "things" (using spatial and numerical descriptions), projections, lengths, perimeters, planes, surfaces, location, etc. Facility with each dimension requires a sense of "benchmarks" and estimation, direct measurement and derived measurement skills. *Shape* is a category describing real images and entities that can be visualized (e.g., houses and buildings, designs in art and craft, safety signs, packaging, snowflakes, knots, crystals, shadows and plants), as well as highly abstract "things" greater than three dimensions. Direction and location are fundamental qualities called upon when reading or sketching maps and diagrams. A basic numeracy task in this fundamental aspect could be shape identification whereas a complex task might involve describing the change in the size of an object when one dimension is changed.

Pattern, Functions and relationships. It is frequently written that mathematics is the study of patterns and relationships. Pattern is seen as a wide-ranging concept that covers patterns encountered all around us, such as those in musical forms, nature, traffic patterns, etc. It is argued by Senechal (1990) that our ability to recognize, interpret, and create patterns is the key to dealing with the world around us. The human capacity for identifying relationships and for thinking analytically undergirds mathematical thinking. Algebra - beyond symbolic manipulation - provides a tool for representing relationships between amounts through the use of tables, graphs, symbols, and words. The ability to generalize and to characterize functions, relationships between variables, is a crucial gateway to understanding even the most basic economic, political or social analyses. A basic level numeracy task might require someone to describe how items are arranged in a package; developing a formula for an electronic spreadsheet would put a higher level of demand on the individual.

Data and chance. Data and chance encompass two related but separate topics. *Data* covers "big ideas" such as variability, sampling, error, or prediction, and related statistical topics such as data collection, data displays, and graphs. Modern society demands that adults interpret and produce organizers of data such as frequency tables, pie charts, graphs and to sort out relevant from irrelevant data. *Chance* covers "big ideas" related to probability, subjective probability, and relevant statistical methods. Few things in the world are 100% certain; thus the ability to attach a number that represents the likelihood of an instance is a valuable tool whether it has to do with the weather, the stock-market, or the decision to board a plane. In this mathematical category, a simple numeracy skill might be the interpretation of a simple pie chart; a more complex task would be to infer the likelihood of an occurrence, such as predicting the weather, based upon past information.

Change. This term describes the mathematics of how the world changes around us. Individual organisms grow, populations vary, prices fluctuate, objects traveling speed up and slow down. Change and rates of change help provide a narration of the world as time marches on. Additive, multiplicative, exponential patterns of change can characterize steady trends; periodic changes suggest cycles and irregular change patterns connect with chaos theory. Describing weight loss compares as a simple task to calculating compounded interest.

3.4 Facet 4: Representations of mathematical information

Mathematical information in an activity or a situation may be available or represented in many forms. It may appear as *concrete objects* to be counted (e.g., people, buildings, cars, etc.) or as *pictures* of such things. It may be conveyed through symbolic notation (e.g., numerals, letters, and operation or relationship signs). Sometimes, mathematical information will be conveyed by *formulas*, which are a model of relationships between entities or variables.

Mathematical information may be encoded in visual displays such as a *diagram* or *chart*; *graphs* and *tables* may be used to display aggregate statistical or quantitative information (by displaying objects, counting data, etc.). Similarly, a *map* of a real entity (e.g., of a city or a project plan) may contain information that can be quantified or mathematized.

Finally, a person may have to extract mathematical information from various types of *texts*, either in prose or in documents with specific formats (such as in tax forms). Two different kinds of text may be encountered in functional numeracy tasks. The first involves mathematical information represented in textual form, i.e., with words or phrases that carry mathematical meaning. Examples are the use of number words (e.g., "five" instead of "5), basic mathematical terms (e.g., fraction, multiplication, percent, average, proportion), or more complex phrases (e.g., "crime rate increased by half") which require interpretation. The second involves cases where mathematical information is expressed in regular notations or symbols (e.g., numbers, plus or minus signs, symbols for units of measure, etc.), but is surrounded by text that despite its non-mathematical nature also has to be interpreted in order to provide additional information and context. An example is a bank deposit slip with some text and instructions in which numbers describing monetary amounts are embedded.

3.5 Facet 5: Other enabling factors and processes

The way in which each person manages his interpretations and responses to the contexts, tasks, and mathematical representations described above as facets of adult numeracy depends of course on his or her mathematical knowledge, whether formally learned, informally developed, or self-invented. This includes but is not limited to the understanding and ability to apply concepts, ideas and procedures detailed in many strands for school curricula or in many school-based assessments under titles such as Whole numbers and basic operations, Ratios, percents, decimals, and fractions, Measurement, Geometry, Algebra, or Probability and statistics. These topics are interwoven into the five areas of "mathematical information" described above in subsection 3.3, and are assessed by the items in the Numeracy scale.

Numerate behavior, however, depends on an integration of mathematical knowledge bases with broader reasoning and problem-solving skills and strategies needed to be able to think and to act mathematically. Further, numerate behavior depends on the integration of the above with the literacy skills, the dispositions (beliefs, attitudes, habits of minds, etc), and prior experiences and practices that an adult brings to each situation. These are briefly discussed below, with comments on the extent to which each is assessed in ALL.

Problem-solving skills. Throughout life, adults develop or apply diverse strategies to manage their quantitative situations. Some strategies or skills may be based on prior formal learning, while others may be self-invented or adapted to fit the situation at hand. To solve many computational problems or to figure a way to manage certain quantitative tasks, people have to re-construct reality in a mathematical way, for example, model or mathematize. They can do so either on their own or in discussion with other people. Problem-solving strategies may include, e.g., extracting relevant information from the task/activity; rewriting/restating the task; drawing pictures, diagrams or sketches; guessing and checking; making a table; and/or generating a concrete model or representation. To some extent, these strategies are appropriate for determining a response to the Numeracy items, but they are assessed more fully in the Problem Solving Scale of the ALL Survey.

Literacy skills. The ability to read, write, and talk are important skills in undertaking a numeracy task or activity or communicating the outcomes of working on such tasks. In cases where "mathematical representations" involve text, one's performance on numeracy tasks will depend not only on formal mathematical or statistical knowledge but also on reading comprehension and literacy skills, reading strategies, and prior literacy experiences. For example, following a computational procedure described in text (such as the instructions for computing shipping charges or adding taxes on an order form) may require special reading strategies, as text is very concise and structured. Likewise, analyzing the mathematical relationships described in words requires specific interpretive skills, e.g., realizing that "four more than" is a different relationship than "four times as much."

Beliefs and attitudes. A person may not necessarily act in numerate ways, even if she or he can demonstrate high ability on a numeracy test. The way in which a person responds to a numeracy task—including overt behavior or actions as well as cognitive processes and the propensity to adopt a critical stance—will depend not only on knowledge and skills but also on beliefs, attitudes, habits of mind, and prior practices. In some cultures, some adults, including highly educated ones, decide that they are not "good with numbers." These sentiments or self-perceptions are usually attributed to negative prior experiences they have had as pupils of mathematics (Tobias, 1993), and stand in contrast to the desired sense of "at-homeness with numbers" (Cockcroft, 1982). Such attitudes and beliefs can interfere with one's motivation to develop new mathematical skills or to tackle math-related tasks, and may also affect test performance (McLeod, 1992).

In realistic contexts, adults with a negative mathematical self-concept may elect to avoid a problem with quantitative elements, address only a portion of it, or prefer to delegate a problem, e.g., by asking a family member or a salesperson for help. Such decisions or actions are indeed the prerogative of a manager and can serve to reduce both mental and emotional load (Gal, 2000). Yet, such actions may fall short of autonomous engagement with the mathematical demands of real-world tasks (as noted in the core definition of numeracy used here), carrying negative consequences, e.g., not being able to fully achieve one's goals. Therefore, prior experiences and existing habits of coping with mathematical and numeracy situations may be influential.

Numeracy-related practices and experiences. Research suggests that, for adults as well as for children, mathematical knowledge develops both in and out of school (e.g., Schliemann & Acioly, 1989; Saxe, 1991; Lave, 1998). Saxe and his colleagues have written about the importance of cultural practice in the development of mathematical thinking and how such practices profoundly influence an individual's cognitive constructions and mathematical ideas, depending, e.g., on the artifacts or tools they use, the nature of the measurement systems in their culture, the counting or calculating devices (abacus, calculator) they use, the distribution of work among family members, or general patterns and types of social activity.

Mathematical experiences and practices, whether at work, home, when shopping, or in other contexts, can be both the result of a certain skill level, or the cause of skill levels. For this reason, it was deemed important to add several items to the Background Questionnaire that examine the frequency of performing certain related numeracy tasks in different contexts including how often artifacts such as calculators or computer spreadsheets are used.

4. Factors affecting complexity of Numeracy Items

Because of the scarcity of research on adults' use and application of numeracy, there is insufficient empirical knowledge to determine what factors make a numeracy activity or task more difficult or complex. One of the more exciting, and challenging, aspects of the project was the development of a scheme to account for the difficulty of different numeracy assessment tasks. We sought such a scheme to inform item development, i.e., help in the creation of items that spread over a range of difficulty levels. However, if the scheme could be shown to correlate with actual difficulty levels of items as measured in actual testing of a sample of individuals, it could also be used to help *explain* observed performance. Given its importance for both item development and interpretation of results, the complexity scheme is described in detail in this section.

4.1 Previous research on task complexity

In IALS, three factors were found to be the principal components of task difficulty regarding literacy or text-based tasks: plausibility of distractors, type of match required, and type of information required. The difficulty of the Quantitative Literacy tasks appeared to be a function of several other factors:

1. The particular arithmetic operation required to complete the task
2. The number of operations needed to perform the task
3. The extent to which the numbers are embedded in printed materials
4. The extent to which an inference must be made to identify the type of operation to be performed (i.e. problem transparency; see below)

The IALS QL difficulty factors overall fit those used in large-scale assessments of mathematical skills (with children), which often make use of three or four factors:

1. *The mathematical concepts involved:* number systems and number sense, spatial and geometrical topics, functions and algebra, chance/statistics topics, etc. Concepts that are related to topics taught in lower grades are considered easier.
2. *The complexity of operations*: addition, subtraction, multiplication, and division, as well as dealing with whole numbers, with decimals, and with percents. Operations that are related to topics taught in lower grades are considered easier.
3. *The number of operations*: one-step problems are considered easier than multi-step problems.
4. *Problem transparency*: This factor is sometimes relevant; it refers to the extent to which the problem situation includes clearly identified numbers or entities and the extent to which it is clear what operations or actions to perform. To the extent that these are not clear or transparent, respondents have to extract needed information by applying comprehension and inference strategies, making the task more complex.

There are other adult-related assessment projects on which to draw to develop the levels of complexity. Both the Essential Skills Research Project and the Applied Numeracy sub-test of the Work Keys test battery (American College Testing, 1997) use a two-factor model of complexity in their description of numeracy levels. The first factor "operations required;" is seemingly straightforward and refers to the difficulty of operations called for. However, this is complicated by the level of difficulty of the numbers being manipulated: computations that include fractions and decimals are usually more difficult than those with whole numbers.

The Essential Skills model spells out two sequences of complexity on this factor: *Operations* and *Translation* of information (sometimes called 'problem transparency').

Operations:

1. Only the simplest operations are required and the operations to be used are clearly specified. Only one type of mathematical operation is used in the task.
2. Only relatively simple operations are required. The specific operations to be performed may not be clearly specified. Tasks involve one or two types of mathematical operation. Few steps of calculations are required.
3. Task may require a combination of operations or multiple-applications of a single operation. Several steps of calculation are required. (More advanced operations may call for multiplication or division.)
4. Tasks involve multiple steps of calculation.
5. Tasks involve multiple steps of calculation. Advanced mathematical techniques may be required (e.g., percents, ratios, proportions).

Translation (Problem Transparency)

1. Only minimal translation is required to turn the task into a mathematical operation. All the information required is provided.
2. Some translation may be required or the numbers needed for the solution may need to be collected from several sources. Simple formulae may be used.
3. Some translation is required but the problem is well defined.
4. Considerable translation is required.
5. Numbers needed for calculations may need to be derived or estimated; approximations may need to be created in cases of uncertainty and ambiguity. Complex formulae, equations or functions may be used.

Two considerations prompted us to question the appropriateness of using mathematics-related frameworks (from Essential Skills or elsewhere) as the sole source for development of a complexity scheme for items assessing *adults*' ability to cope with real-world numeracy tasks. First, effective coping with many real-world quantitative problems depends upon people's ability to make sense of and interact with different types of texts. This is hardly recognized by the Essential Skills model. Hence, it was essential to add difficulty factors that acknowledge the inherent links between literacy and numeracy, quite similar to those used in IALS.

Another, albeit a more restricted consideration, is that the ordering of complexity of tasks by the type of operation performed may not be as clear with adults as it may be with children. Such ordering in school-based assessments is predicated on traditional school curricula, where more advanced topics are learned at higher grades. However, adults are known to use a lot of invented strategies, perhaps more so, and more efficiently so, than children. Multiplication or division problems, which can prove relatively hard for some young people, may be solved by seemingly simpler strategies, such as by repeated addition or repeated subtraction; complex numbers may be broken down in ways that ease mental load, and so forth. In addition, adults' familiarity with everyday contexts, such as with monetary entities, facilitates their performance with some seemingly advanced concepts. For example, specific benchmark values of fractions and percents, such as 1/2, 1/4, 50%, or 25%, are familiar to many people; as a result, they may be easier to manage than expected, violating curriculum-based ordering of difficulty. Hence, an *overall complexity level* has to be used, in order to weight these "inconsistencies" in ordering of difficulty levels proposed in other schemes.

4.2 Complexity factors in the ALL survey

The above literature review suggests that a framework of factors affecting the complexity of numeracy tasks should not only address factors related to the numerical and textual aspects of tasks, but should also address other issues. It should treat separately the number of operations and the type of operations from the *type of mathematical (or statistical) information to be processed*, which may involve numbers explicitly but also other types of mathematical information. In so doing, the desired framework of complexity factors should take into account the broad scope of the definition of numeracy, i.e., reflect the variation within contexts, the range of mathematical ideas/content, the types of possible responses, and the types of representations that cut across adult life contexts.

With the above considerations in mind, five key factors have been identified that are predicted to affect, separately and in interaction, the difficulty level of numeracy tasks to be used in the ALL survey. These five "complexity factors" are outlined in Table 2 and are organized in two sets: two factors that address mainly textual aspects of tasks, and three factors that address the mathematical aspects of tasks. These five factors are listed separately for clarity of presentation, but in actuality are *not* independent of each other and do interact in complex ways. Each factor is examined in some detail below, followed by a later subsection that describes the calculation of an overall complexity level for each item, taking into account all five factors.

Table 2:

Complexity factors—Overview

Aspects	Category	Range
Textual aspects	1. Type of match/problem transparency	Obvious/explicit to embedded/hidden
	2. Plausibility of distractors	No distractors to several distractors
Mathematical aspects	3. Complexity of Mathematical information/data	Concrete/simple to abstract/complex
	4. Type of operation/skill	Simple to complex
	5. Expected number of operations	One to many

Type of Match/Problem Transparency. This is a combination of the factor of Problem Transparency outlined above, and of an IALS factor called Type of Match. Problem Transparency is a function of how well the mathematical information and tasks are specified and includes aspects such as how apparently the procedure is set out, how explicitly the values are stated, etc. Type of Match refers to the process that a respondent has to use to relate the requested action in the question to the information in the task or text, which can range from a simple action of locating or matching to more complex actions that require the respondent to perform a number of searches through the information given. This measure of complexity for a numeracy task incorporates the degree of text embeddedness of the mathematical information.

In easy tasks, the type of information (e.g., numerical values) and the operations needed are apparent and obvious from the way the situation is organized. In more difficult ones, the values must be located or derived from other values; the operations needed may have to be discovered by the performer, depending on his or her interpretation of the context and of the kind of response expected. As well, numeracy situations may involve text to varying degrees, and this text may be of different degrees of importance. There may be a situation where there is little or no text. Some situations may involve pure quantitative information that is to be interpreted or acted upon with virtually no text or linguistic input. In other words, the performer derives all the information needed to respond from the objects present in the situation or from direct numerical displays.

At a higher level, some textual or verbal information may be present alongside the mathematical information. The text can provide background information about the problem situation, or some instructions. For example, a bus schedule, cooking instructions, and a typical school-type word problem all involve some text and some numbers. Still other situations would be heavily text-based or may not involve any numbers or mathematical symbols at all, just plain text. The task will contain mathematical or statistical information that a person needs to understand and, in some cases, act upon, but it will be much less transparent. It may be heavily embedded in dense text or may require using information from a number of sources within or even outside the text/task.

This factor requires that a task will be analyzed in terms of the questions: *How difficult is it to identify and decide what action to take?*, and *How many literacy skills are required?*

Plausibility of Distractors. This variable is literacy related, even though it can involve mathematical components. In general, literacy tasks are easiest to process when there are no plausible distractors in the text, that is, there is no other information in the text that meets any of the requirements of the task. At higher levels of difficulty, tasks can involve irrelevant information both within the question as well as within the text. In terms of mathematical information, a low level of plausible distractors would mean that no other mathematical information was present apart from that requested, making the numbers or data required easy to identify. At a higher level, there may be either some other mathematical information in the task (or its text) that could be a distractor, or the mathematical information given or requested could occur in more than one place. A higher level of complexity could also mean that outside information (e.g. the knowledge of a formula) may be needed to answer the question.

This factor requires that a task will be analyzed in terms of the questions: *How many other pieces of mathematical information are present?*, and *Is all the necessary information there?*

Complexity of Mathematical Information. Some situations present a person with simple mathematical information, such as concrete objects (to be counted), simple whole numbers, or simple shapes or graphs. At lower skill levels, the information will be more familiar, whereas at higher levels, the information may be less familiar. Situations will be more difficult to manage if they involve more abstract or complex information, such as very large or very small numbers, unfamiliar decimals or percents, information about rates, or dense visual information, as in a diagram or complex table.

This factor requires that a task will be analyzed in terms of the question: *How complex is the mathematical information that needs to be manipulated or managed?*

Type of Operation/Skill. Some situations require simple operations, such as addition or subtraction, or simple measurement (e.g., finding the length of a shelf), or recognition of shape. These are usually easier to analyze mathematically than situations that require multiplication or division, and than situations that require using exponents. While the difficulty of recognizing and carrying out the operation implied by a situation (be it additive, multiplicative, etc.) has direct bearing on task complexity, there may be exceptions that occur when alternative approaches are obvious. There are some tasks that combine both interpretive and generative skills and may involve a deeper conceptual understanding than merely carrying out a procedure. Other more complex tasks may involve an explanation of one's reasoning. The interpretation of information appearing in graphs, for example, becomes more complex if comparisons, conjecturing, or "reading beyond the information given" is required.

This factor requires that a task will be analyzed in terms of the question: *How complex is the mathematical action that is required?*

Expected Number of Operations. Tasks that require acting upon the mathematical information given may call for one application (step) of an operation, or for one action (e.g., literal reading of information in a table, or measurement). More complex tasks will demand more than one operation, which may be the same or similar to one another, such as the steps involved in multiple passes on the data or text. Still more complex tasks are those that involve the integration of several different operations.

This factor requires that a task will be analyzed in terms of the question: *How many steps and types of steps are required?*

4.3 Overall complexity level

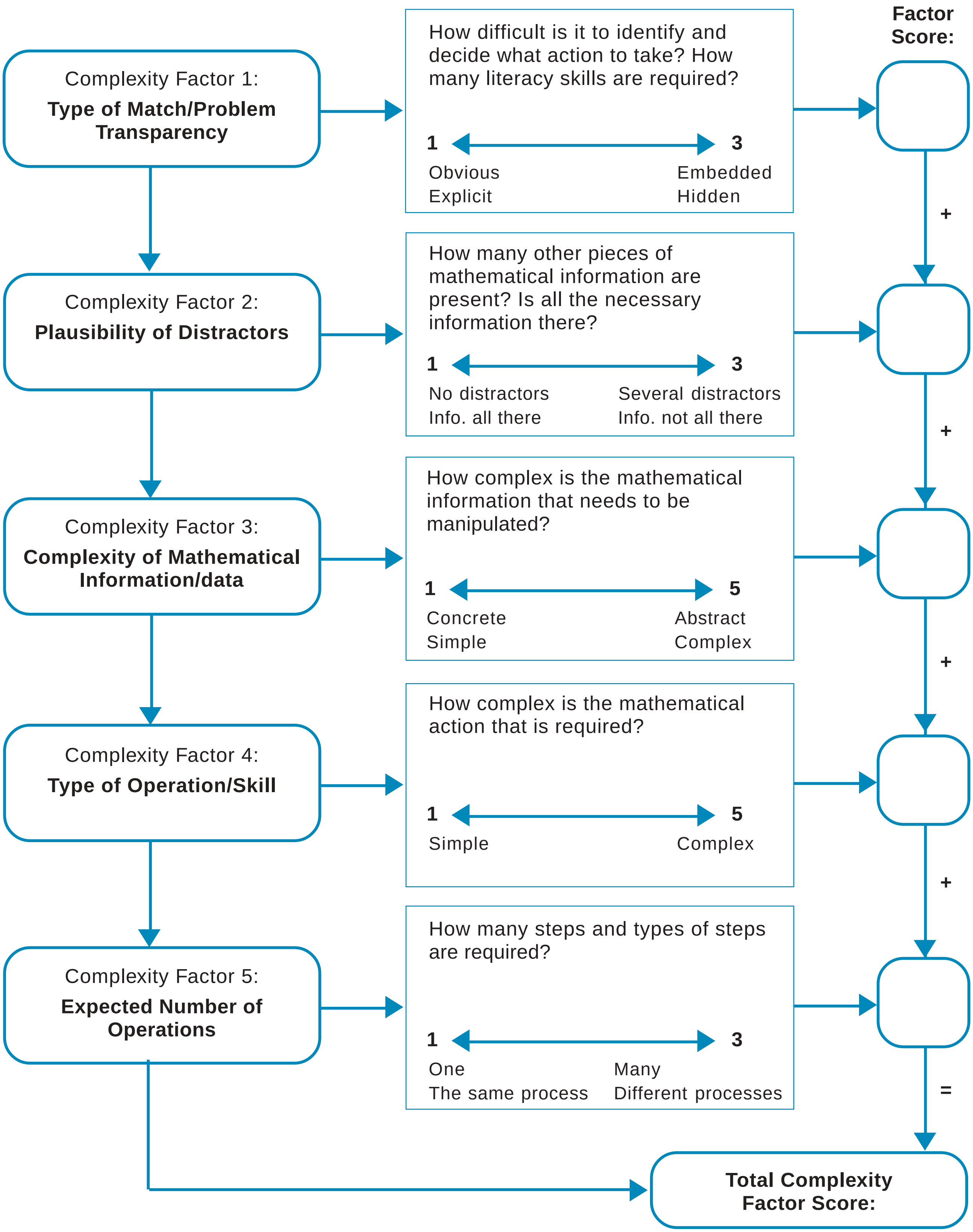
It is possible to estimate the overall difficulty level of a specific item by first scoring the item on each of the five factors of complexity, according to the levels described in Appendix 2, and then summing together the scores for each factor. Figure 1 below explains the process; Appendix 2 describes each level of the five factors in detail. The total summary score can range between 5 (easiest) and 19 (most difficult).

The estimation process outlined in Figure 1 suggests that each factor has a separate contribution to an item's overall difficulty or complexity. However, it can be hypothesized that as tasks become more complex, actual performance on items may increasingly depend not only on each factor by itself, but also on the interplay or interaction between them. Hence, the computational process suggested in Figure 1 can provide only approximate information about an item's anticipated difficulty level.

Further, the difficulty of a task cannot in some cases be predicted without taking into account characteristics of the person who interacts with the task. The same task may be more difficult for some individuals and less difficult for other individuals, depending on factors such as their familiarity with the context in which a task is situated, knowledge of formal mathematical notations, background world knowledge, as well as general literacy, problem-solving, and reasoning skills. For example, it could be predicted that a task that involves the composition of a fertilizer would be more difficult for an urban apartment dweller than for a rural farmer whereas a task that uses a bus schedule would be more difficult for the farmer. For the above reasons, the prediction of the difficulty of a task in isolation of detailed knowledge about the respondent himself can only be an estimate.

Figure 1

Complexity Flow chart



Despite the above limitations, the scheme of complexity factors developed for numeracy assessment in ALL comprises a theoretical contribution. It provides a conceptual basis for predicting the different levels of complexity of a broader range of items well beyond those involving arithmetic operations only. To the extent these predicted difficulty levels later prove to correlate with actual difficulty (such as in terms of percent correct on different items), this scheme can also assist in interpreting survey results, as it can help to explain *why* some items are harder than others and what kinds of cognitive processes are called for by different tasks.

Indeed, results from the feasibility and pilot studies described later in this report show that predicted difficulty and observed difficulty were highly correlated. Nonetheless, further distillation and validation are needed, and the scheme in its current state is included mainly to show the logic behind a process used during item development.

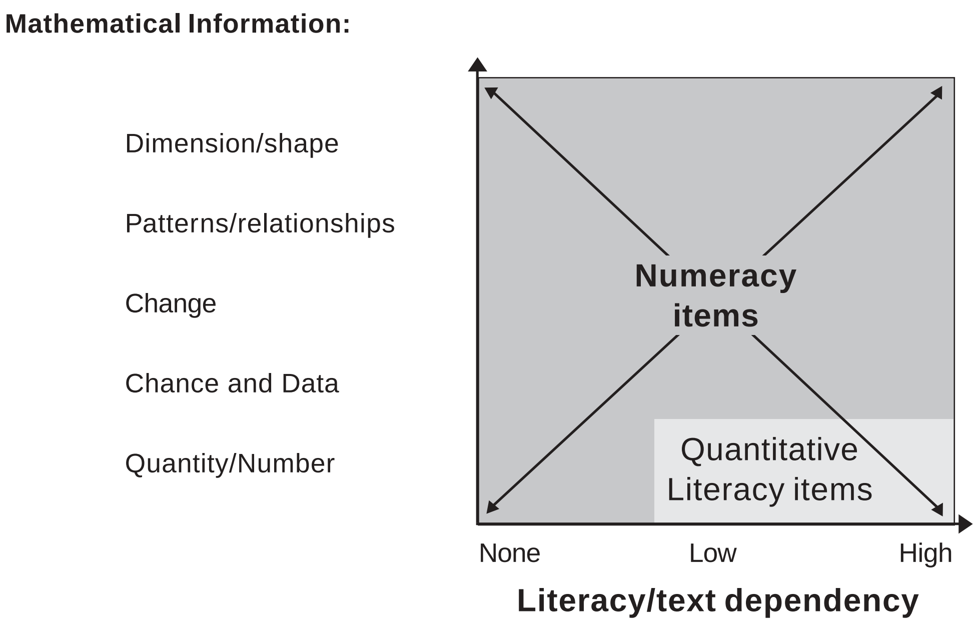
5. Summary: Directions and challenges in assessing Numeracy in ALL

Given the increasing need for adults to continuously adapt to changing citizenship and workplace demands (European Commission's White Paper, 1996), it is vital that nations have information about their workers' and citizens' numeracy in order to understand skill distributions in the population as a whole as well as in specific groups, and be able to plan effective lifelong learning opportunities.

Numeracy, a domain to be assessed in the ALL survey, has been conceptualized in this report as a broader construct than Quantitative Literacy that was assessed in the earlier IALS. Numeracy requires more varied responses (order, count, estimate, compute, measure, interpret, explain) to a wider range of mathematical information (quantity, dimension and shape, pattern and relationships, data and chance, and change) that may be embedded in text in varying degrees. Figure 2 illustrates the difference between the scope of topics that should be encompassed by Numeracy and Quantitative Literacy items.

Figure 2

Numeracy versus Quantitative Literacy



Ideally a survey of adult numeracy skills would ask respondents to complete tasks that are couched in real-life situations and that encompass all components of the facets of numerate behavior described earlier. Further, respondents would ideally respond to in-depth and qualitative interviews in order to understand their answers and reactions as well as their thought processes. Since the ALL is a paper-based household survey, however, there are restrictions on the 'real' nature of the situations that could be included. In addition, the available options for presenting questions and for eliciting, recording, and reliably scoring responses are limited when large samples across wide ranging populations have to be surveyed. Finally, given that numeracy involves a sizable number of facets and subcomponents, practical limitations on the number of items that can be administered to respondents to assess each of the domains covered in ALL mean that not all aspects of the facets of numerate behavior can necessarily be covered at an equal depth.

Operating within these constraints, the goal for the item development process described in the next part was to create Numeracy items that are based on realistic stimuli and include tasks that serve a purpose for most people within their culture. A scheme of factors affecting the complexity of numeracy tasks was developed as a tool to support item development and to help explain performance on Numeracy items. Information derived from performance on the Numeracy scale was designed to be supplemented by information gathered via the Background Questionnaire about prior practices and attitudes related to numeracy, in addition to other important explanatory variables such as those related to demographic data and literacy practices. The process of developing and evaluating items along with their scoring rubrics is described in Part B.

Closing this summary of challenges in converting the conceptual framework of Numeracy into an assessment tool, it must be noted that Numeracy cannot be fully separated from Document Literacy and Problem Solving, two other important domains assessed by ALL. Numeracy is somewhat related to Document Literacy, due to the role that text and graphical or tabular displays play in both constructs. Likewise, Numeracy is not fully distinguishable from Problem Solving, as certain numeracy tasks require the implementation of problem solving processes, albeit brief at times, while in contrast certain “problems” (e.g., designing a budget) involve numerical content.

Hence, a final challenge in designing items for the Numeracy scale was to achieve a sensible “division of labor”. On one hand, it is important to prevent excessive overlap between what is measured by the Numeracy scale and the type of skills assessed by scales in other domains. On the other hand it is also important to maintain a reasonable coverage of the core issues encompassed by the construct of numeracy as envisioned in this report, even if this creates some acceptable degree of overlap with other scales. Such overlap, we argue, is not a limitation of the conceptual framework developed to describe Numeracy and numerate behavior, but simply reflects the complex nature of real-world situations and tasks. While it is convenient for educators and designers of assessment tools to think of skill areas as distinct and having clear-cut boundaries, real-life situations may require that people who have to manage and respond to them rely on interrelated knowledge and skills.

Part B

Item development

6. Approach

Throughout the item-production stages, the Numeracy team used the principles or guidelines outlined below as the basis for creating items. (Note: Appendix 1 includes sample items illustrating the type of stimuli and questions developed).

a. Items should cover as many aspects within each of the four facets of the numeracy construct as possible. Items were generated so as to require the activation of a broad range of skills and knowledge included in the construct of numeracy, as portrayed in the conceptual framework described in Part A and depicted in Table 1.

b. Items should include a realistic stimulus and one or more questions that adults might be expected to encounter, given such a stimulus. The aim was for the item content to be familiar and the questions purposeful to respondents across cultures. Most tasks were derived from real-life stimuli, from situations that are easily understood or that can be expected to be of importance or relevance in different cultures.

c. Items should have a free-response format. Items were structured to include a stimulus (e.g., a picture, drawing, display) and one or more questions, the answers to which the respondent writes in his own words.

d. Items should spread over different levels of difficulty. Items were produced to span the range of difficulty levels covered by the theoretical complexity factors outlined earlier. Attention was paid to generating some items at the lowest level of complexity, which are of interest in countries where policies and educational programs may be earmarked for low-skill populations.

e. Items should vary in the degree to which the task is embedded in text. While about 1/3 of items were embedded in or included relatively rich texts, about 2/3 of the items were designed to use little or no text, to reduce overlap with the Document and Prose literacy scales as they existed in IALS and continue in ALL.

f. Tasks should be relatively short. The use of short tasks enabled the inclusion of a large number of diverse stimuli and questions, thus allowing for coverage of many key facets of the numeracy framework. Items were not designed to simulate extended problem-solving processes, thus reducing overlap with the Problem-Solving scale.

g. Items should overcome issues associated with different unit systems. Items were designed so that different currency systems as well as different systems of measurement (metric or Imperial) could be applied to the numbers or figures used. Such items can retain equivalency with respect to their mathematical or cognitive demands and appear familiar to different populations even after being translated.

7. Production and evaluation of items

The creation of items for the Numeracy assessment progressed through three stages: Two stages involving production of items and their testing in two countries on relatively small samples, and the third stage involved a much larger Pilot testing process.

7.1 Stage 1 (1998-1999): Production and field-testing of a first item pool

Based on the above general principles, a pool of over 80 items was generated by team members, based on their experience in research, assessment, and teaching with both school-based and diverse adult and workplace learner populations in several countries.

Production grid. Items were created so as to fill cells within an item production grid with four key dimensions that match the conceptual facets outlined in Table 1:

1. Type of purpose / context: *everyday, societal, work, further learning*.
2. Type of response: *identifying or locating, acting upon* (*order/sort, count, estimate, compute, measure, model) interpreting, communicating about*.
3. Type of mathematical or statistical information: *quantity, dimension, patterns/relations, data/chance, change*. The content of the tasks was also conceived, however, in terms of common school-based mathematics topics more familiar to policy makers and educators, i.e., whole numbers and basic operations; ratios, percents, decimals and fractions; measurement; geometry; algebra; and statistics.
4. Type of representation of mathematical or statistical information: *numbers, formulae, pictures, diagrams, graphs, tables, texts*.

Scoring. Guidelines for scoring responses were designed to classify them into three general groups: "correct", "any other response" (i.e., wrong answers) and "not attempted" (i.e., no indication the respondent tried an item). However, for many items, *multiple* codes were prepared to capture different types of "correct" or "wrong" answers and thus enable an analysis of error patterns and shed light on the extent to which instructions are understood and items elicit the expected type of responses. In some items that require estimation or measurement, multiple codes were prepared to capture responses that may have different degrees of accuracy yet still fall within a "correct" or "wrong" region, in order to understand the level of accuracy that respondents adopt.

Non-cognitive items. Research literature suggests that the way in which a person responds to a numeracy task, including overt actions as well as internal thought processes and the adoption of a critical stance, depend not only on knowledge and skills but also on negative attitudes towards mathematics, beliefs about one's mathematical skills, habits of mind, and prior experiences involving tasks with mathematical content (Cockcroft, 1982; Lave, 1988; Schliemann and Acioly, 1989; Saxe, 1991; McLeod, 1992; Gal, 2000). Hence, the Numeracy team also prepared several scales for the Background Questionnaire, with questions designed to measure numeracy practices at home and at work, attitudes and beliefs about mathematics, and information about the environment in which the respondent learned mathematics while in school. Such scales may help in explaining performance on numeracy tasks, as well as understanding respondents' status on variables of interest to policy makers, such as participation in further learning or employment status.

First feasibility study. Of the items generated at this stage, 80 items were tested in a feasibility study conducted in the USA and The Netherlands, on samples of about

N=300 per country; each item was answered by about 150 respondents. Results enabled analysis of error patterns and gender bias on items, and assessment of psychometric properties of items in terms of both classical test theory and Item Response Theory (IRT) parameters. Comments made by respondents in focus groups suggested that only a few items were at times misunderstood. Statistical analyses showed that most items have adequate psychometric properties, are answered roughly in the same way by males and females, and that the items tested cover a wide range of difficulty levels. (Four items from the QL scale used in IALS were also tested to enable rough calibration of the item difficulty estimates obtained on the basis of the feasibility sample, in light of the difficulty levels of these four items in the much larger and nationally representative samples used in IALS).

The feasibility study enabled evaluation of several other important issues that arise during item translation and adaptation to different cultural contexts, e.g., in terms of using different monetary, length, or volume units. Results showed that performance on most items is comparable across languages even when they use different units of measurement. Also, responses were scored independently by two scorers, and an analyses of scorer agreement showed very adequate scoring reliability, suggesting that the scoring schemes and scoring instructions were well understood to scorers operating in two languages.

Further analyses revealed interesting patterns of correlations between respondents scores on some of the non-cognitive scales and their overall performance on Numeracy items. This suggested that some of the attitudinal and belief items can indeed be used to help in understanding cognitive performance.

Overall, 68 items out of the 80 items tested appeared to satisfy all selection criteria. For these 68 items, it was found that item difficulty (in terms of percent correct on an item) was highly correlated ($r = -0.793$) with predicted item complexity as determined during item development based on the factors outlined in Figure 2 and detailed in Appendix 2. This indicated that the complexity scheme could serve as a useful aid to inform the direction in which features of new items should be varied so as to reach a pre-determined or desired level of difficulty. The detailed scheme included in the appendix is a refinement of the original and continues to represent work in progress. Because of the recursive nature of the testing of this scheme (e.g., the same individuals wrote the scheme and rated the complexity of items), caution should be exercised in further interpretive use of the present version; further refinement and validation work is necessary.

External review. As part of work at this stage, the conceptual framework and some sample items were sent for review and comment by a panel of 16 experts from 9 countries. The reviews highlighted that there is a range of conceptions regarding the terrain covered by the term “numeracy”, as expected in light of the conceptual analysis summarized in Part A. Nonetheless, the reviews overall supported the conceptual framework for Numeracy assessment developed for the ALL project, and endorsed the approach to item production described earlier in this section.

7.2 Stage 2 (1999-2000): Additional item production, second feasibility study

Following the successful completion of Stage 1, additional items were contributed by experts from four countries: Austria, the Czech Republic, Hungary, and Sweden, in order to enrich the pool of Numeracy items and increase its cultural diversity. Some of these items were selected and adapted by the Numeracy team in order to fit the item production grid and the item development principles.

A small-scale feasibility study was conducted in the USA and The Netherlands, in which 44 additional items were tested on samples of about 55 cases in each country. Statistical analyses conducted at this stage were quite similar to those used during the first feasibility study, although the smaller sample prevented the use of IRT analysis.

Out of all the items tested in the two feasibility studies conducted in Stage 1 and Stage 2, a pool of 81 items was selected for further testing at the Pilot stage. These items had adequate psychometric characteristics, covered diverse levels of difficulty, encompassed key facets of the conceptual framework for Numeracy, and could be adapted without difficulty to a language other than English and to different units of measurement. These items did not show any remarkable error patterns indicating that stimuli, questions or task contexts were misunderstood. In addition, two easy Numeracy items were selected to become part of the "Core", the screener test that is to be successfully passed by each respondent in order for him or her to receive one or more full-length test booklets with items assessing performance in the ALL skill domains.

7.3 Stage 3 (2001-2003): Preparations for pilot testing in participating countries

Scoring materials. Once items for the Pilot stage were selected, the Numeracy team prepared detailed scoring guidelines for each item, as well as a training manual for scorers. In addition, the team prepared a detailed manual describing "critical elements" of each item that should be kept constant during the translation and adaptation work that came next.

Translation and adaptation for Pilot. The 81 items selected at the end of Stage 2 were translated and adapted by all participating countries, sometimes into multiple languages within the same country (e.g., Canada created English and French versions, and Switzerland created German, French, and Italian versions). The translation process was supported by training workshops held in 2000 and 2001 and by the support materials prepared by the Numeracy team, i.e., the manual describing critical elements of each item that should be kept constant across language versions, and a "translation and adaptation manual". Item adaptation aimed to maintain cognitive equivalency in terms of their task demands. Thus, for example, when units of measure or monetary values in some computational items were adapted to various country situations, guidelines emphasized the need to keep item demands comparable.

Further, each country prepared not only country-specific and language-specific sets of items, stimuli, and response pages, but also adapted the scoring instructions and the training manual for scorers. During this stage, all translators could also post questions by e-mail to a hotline staffed by members of the Numeracy team.

The next section provides an overview of the results from the Pilot phase and further explanations regarding the properties of the final 40 items selected for the Main assessment.

Part C

Pilot Study: Summary of Numeracy-related results and decisions

8. Overview and purposes of the pilot stage

The ALL Pilot study was conducted in 2002 in six countries (and five languages): Belgium (French), Canada (French and English), Italy, Norway, Switzerland (German, French, Italian), and the United States. Large samples responded to test booklets assessing the key skill domains targeted by ALL: Document literacy, Prose literacy, Numeracy, and Problem-solving. The Background Questionnaire (BQ) provided additional information about respondents' experiences with Information Technology and about numerous variables of importance such as employment status, consumption of health services, participation in learning activities, and perceptions of quality of life. The BQ also collected additional personal information that could help to explain or interpret results pertaining to the key variables.

8.1 General goals

The Pilot study was designed to accomplish two goals:

a. Enable the ALL management and the development teams to evaluate the psychometric characteristics of the item pools created to assess the different skill domains, and provide sufficient data from which to select the items to be included in the test booklets and the BQ that will be used in the Main ALL survey.

b. Enable participating countries to field-test and evaluate all survey materials and administrative procedures planned for the Main ALL survey, such as contacting respondents, setting home visits, test administration, scoring of responses, data capture, and quality assurance mechanisms.

8.2 Numeracy-specific goals

In addition to the two general goals listed above, the Pilot study aimed to provide information regarding specific issues relevant to Numeracy assessment in ALL:

- Enable selection of 40 items out of the 81 Numeracy items tested in the Pilot, whose development was described in Part B. The target number of 40 items was determined on the basis of knowing average response times per item in the feasibility studies, and the overall time expected to be available for Numeracy assessment in the Main study.
- Analyze response patterns and verify that errors are not caused by unclear or problematic stimuli, questions, or instructions, or by inconsistent translation. Also, determine whether the multiple scoring codes created for some items should be maintained for the Main study

- Collect data on the extent to which respondents used the calculator provided as part of the assessment.
- Evaluate the relationships between Numeracy scale scores and variables measured by the BQ, including scores on scales related to numeracy practices and to beliefs and attitudes.

9. Methods

9.1 Respondents

In the ALL Pilot study, respondents aged 16 to 65+ were individually tested in their homes for approximately 90-120 minutes. In each participating country, around 1000-1500 individuals were tested overall, although each respondent was not tested in all of the domains in ALL. Given the goals of the Pilot study, respondents did not comprise a probability sample from each country, although care was taken to recruit individuals from diverse locations and stratify and balance the sample in terms of gender, age groups, and educational levels.

9.2 Procedure and instruments

All respondents were first given a short "Core", a screener test consisting of four simple Prose and Document Literacy tasks and two simple Numeracy tasks, which were read aloud by the examiner. Respondents were then given test booklets that included a subset of items from one or two of the key skill domains, as well as BQ questions. Respondents wrote their answers inside the test booklet. Those who received Numeracy items were free at all times to use a calculator and a ruler provided by the examiner.

Each respondent who received a Numeracy test answered only about half of the 81 items evaluated in the Pilot study. These items were divided into four blocks that were paired in various permutations into four booklets, thus each respondent received two of four blocks. This arrangement enabled testing of the complete Numeracy item pool and evaluation of the relationships between all items, although each respondent answered only half of the items. The total number that tried each Numeracy item averaged about 950 respondents across the countries and languages listed above.

9.3 Scoring

Items were scored by trained teams in each country according to the scoring rubrics and guidelines designed by the Numeracy team. Double-scoring procedures were implemented on a sample of test booklets in each country, to evaluate consistency of scoring and detect problems with scoring instructions. A third scorer arbitrated disagreements. During the scoring process, a member of the Numeracy team answered the questions from scorers in the participating countries on an active electronic listserve.

9.4 Data-analysis

Various statistical analyses were conducted on the available data, with scaling based on Item Response Theory, analysis of difficulty levels of items and of scale characteristics using classical test theory, analysis of the frequency of different types of errors across countries, and correlational analysis of linkages between summary scores on Numeracy items and BQ variables and scales of interest. Key analyses were performed at the Educational Testing Service in Princeton, and some were conducted by members of the Numeracy team and by Statistics Canada.

10. Results

10.1 Items

With the ultimate goal of selecting 40 items for the Main Survey, the Numeracy team first looked at the Pilot data to identify items that may be problematic in terms of psychometric anomalies or scoring unreliability. Some of the key patterns or data examined were the following:

- Wide variation in performance on an item from country to country or across gender groups could mean that the context of the item was not universal or it could reveal discrepancies in either adaptation of items or test administration.
- When the observed performance on an item (for example, "percent correct" in classical test theory terms) deviates significantly from the difficulty level predicted from the theoretical complexity scheme (See Part A), it could indicate misunderstanding of the question or the presence of unexpected factors that cause response errors in some but not all countries.
- A large number of disagreements between scorers for an item could indicate that a scoring rubric was not discriminating properly. In addition, anecdotal reports from scorers on the listserv can be used to flag a few items whose scoring rubrics were difficult to score for various reasons.

Based on such considerations, the team rejected very few of the 81 items, as gross problems were mostly already eliminated after the two feasibility studies and translation and adaptation process. A few problems with specific items were revealed and addressed by technical recommendations. For example, instructions to use the ruler on certain items were made more explicit both for the respondent and the examiner. Recommendations were made on how to standardize the production of stimuli which required that respondents perform a measurement of length but which were not printed to the same exact measurements in different countries. In addition, the complex scoring rubrics developed and used up to the Pilot study for error analysis were collapsed into a simpler correct/incorrect classification. This approach was chosen in light of the advantages in terms of simplifying scoring processes and scorer decisions and after analyses showed no significant loss in the information provided about respondents' skills for the purposes of this survey.

10.2 Scale

The pilot results reaffirmed that the Numeracy items as a whole constitute a cohesive scale. Across all the countries, the mean Chronbach's Alpha coefficient, a measure of the internal consistency of a group of items intended to represent an underlying construct, averaged 0.88 across the four blocks of Numeracy items. A result above 0.80 generally indicates an underlying consistency in a scale designed to measure cognitive skills. An average Alpha of 0.88 is especially informative, considering that the Numeracy items encompass a range of facets of content, and vary in terms of their contexts, literacy demands and response requirements.

Additionally, the construct as defined was validated by the strong correlation between the predicted and observed difficulty of the items in the pilot. With a wider and more diverse population than the feasibility study, the pilot study empirical results (percent correct) were highly correlated (r = -0.799) with the theoretical predictions of difficulty determined with the complexity scheme (Figure 2 and Appendix 2). Again, the reader is advised that further validation and refinement of the detailed scheme is planned.

11. Selection of items for the Main Numeracy assessment

Following the above analyses, 40 items were chosen that together satisfy several requirements or goals that are outlined below. Information about the extent to which the selected 40 items satisfy these requirements is briefly described.

11.1 Facet coverage in scale

A primary goal for the items selected for the Main study is that they represent the various aspects of four facets outlined as part of the numeracy construct, so that the scale overall offers respondents an opportunity to demonstrate a variety of numeracy skills, tested within a range of realistic contexts.

Mathematical Idea and Type of Response. The first focus in selecting items was on covering the two key facets of "mathematical idea" (Facet 3) and "type of response" (Facet 2). A desired mix of the categories within these two facets was created and is listed in Table 3, and items of varying levels of difficulty were chosen to fit within that mix.

It should be noted that the categories within the facets are not mutually exclusive; that is, an item whose primary task involves "dimension" may also require dealing with "quantity" and one that requires interpretation may also involve estimation. Hence, the percentages in Table 3 are not absolute, as items could be classified in more than one way, depending on which of their several key demands are taken into consideration.

Table 3

Expected distribution of items within facet categories

Content area	%	Type of Response	%
Quantity	30	Interpret	35
Dimension	30	Compute	30
Relations	10	Estimate	10
Data	20	Order	10
Change	10	Measure/Count	10
		Model	5

Contexts. Another consideration in selecting items for the Main study was to achieve a desirable mix of the contexts from which the tasks are drawn (Facet 1 in Table 1). Nearly half, 45%, of the 40 items chosen are situated in Everyday Life contexts, 25% are from situations that represent participation in the Larger Society, 20% are from the world of Work, and 10% represent more formal tasks that may be helpful in Further Learning.

Representations. A final concern was to vary the types of stimuli that were included in the final scale. Each type of stimuli listed in the description (Facet 4 in Table 1) is present in the 40 items, i.e., pictures, numbers and symbols, formulae, diagrams and maps, graphs, tables, and, of course, text. With respect to text, the 40 items selected include stimuli (and questions) that span the range from minimum text dependency to higher levels of dependency, as shown in Table 4. The level of literacy/ text dependence of the stimulus and the question is an important consideration, especially when comparisons are made to the QL scale used in IALS, where stimuli were often quite heavily text-based and required considerable literacy skills.

Table 4

Text aspects of items

Text dependency	Percentage of items
Low	37.5%
Medium	35.0%
High	27.5%

11.2 Difficulty levels in scale

A major goal of the ALL Main Survey is to profile the skill distribution of a country's population. The 42 items to be chosen to assess Numeracy in ALL (including the two Numeracy items in the Core screener) should represent a range of levels of difficulty, so that they can discriminate reliably between performance levels of respondents.

Table 5

Distribution of difficulty levels

Level	Number of items
1 (easy)	6
2	10
3	17
4	6
5 (difficult)	3

Tentative estimates of the performance level required of respondents to answer items in nationally representative samples in the Main study were derived on the basis of IRT analyses of the Pilot data and comparisons to data from the Prior IALS, and appear in Table 5. (While IALS and other large-scale comparative surveys of adults' skills represent item difficulty on a continuous scale, when reporting key results, they group items in terms of five difficulty levels, where level 1 refers to easy items and level 5 refers to difficult items). The 40 items (plus two Core items) selected for the Main appear to form a satisfactory distribution across five performance levels. The distribution of items in terms of difficulty levels provides the most information at the center of the expected population distribution, and thus promises a rich field of data from which to draw a profile of the population's numeracy.

11.3 Balance between blocks

Within the structure of the Main Survey, Numeracy items are organized by blocks. The BIB ("Balanced Incomplete Block") design adopted for the ALL survey, which was also used in IALS and other large-scale surveys, dictates that a block of Numeracy items be paired in booklets with a block of Prose and Document Literacy items or with a second block of Numeracy items. Average time requirements recorded during the pilot indicated that the allotment of 20 Numeracy items per block worked well and should be continued for the Main survey.

Hence, another requirement from the 40 items chosen for the Numeracy scale is that they can be divided into two blocks that reflect the difficulty level and the facet coverage of the entire scale, for psychometric as well as for "face validity" reasons. Subsequently, two 20-item blocks were created which are well balanced with respect to the means of the IRT indicators of discrimination (Mean Rbiserial) and difficulty

(Mean b), and in terms of the number of items in each block that represent the five content areas listed in the conceptual framework of numeracy facets, as shown in Table 6 and Table 7, respectively. Thus, the two blocks that make up the Numeracy scale overall satisfy the requirement for psychometric equivalency and each of them represents the richness of the full construct.

Table 6

Blocks by psychometric indicators

	Block A	Block B
Discrimination (Mean Rbiserial)	0.63	0.67
Difficulty (Mean b)	-0.33	-0.33

Table 7

Blocks by content categories

	Block A	Block B
Quantity	6	6
Dimension	7	6
Relations	2	2
Data	3	4
Change	2	2

12. Summary

The results from the Pilot Study provide support both for the content validity and the construct validity of the Numeracy scale. The 40 items selected for the Main Study cover the Numeracy domain as envisioned. They reflect the facets contained in the definition of measurable numerate behavior, include items at all levels of difficulty, and satisfy psychometric requirements common in international comparative studies.

In addition, analysis of the pilot data suggest that "non-cognitive" items, including several item clusters developed by the Numeracy team, can serve as useful covariates of observed performance on Numeracy items and of numeracy skill levels. Some of these items ask about general attitudes concerning mathematical tasks and about recollections of mathematics learned in school settings. Other items ask about a respondent's involvement with particular numeracy tasks on the job, about self perceptions as to whether the respondent's math skills are sufficient to accomplish what is required at work, and examine certain beliefs, attitudes, and practices related to numeracy. Another item asks for a self-assessment of the extent to which a respondent used a calculator during the Numeracy block.

Overall, the forty items chosen for Numeracy assessment can be used to describe the distribution of numeracy skills in the population as well as in important sub-groups. In addition, the supplemental non-cognitive items included in the Background Questionnaire can shed light on possible antecedents and consequences of the numeracy skills that are of interest to policy makers and other audiences.

References

American College Testing. (1997). *Work keys: Preliminary technical handbook*. Iowa City, IA: Author.

Australian Education Council (The Mayer Committee). (1992). *The key competencies report: Putting general education to work.* Canberra: The Australian Education Council and Ministers for Vocational Education, Employment, and Training.

Baker, D., and Street, B. (1994). Literacy and numeracy: Concepts and definitions. In T. Husen and E.A. Postlethwaite (Eds.), *Encyclopedia of education*. New York: Pergamon Press.

Beazley, K. (1984). Education in Western Australia: Report of the Committee of Inquiry into Education in Western Australia. Education Department of Western Australia.

Benn, Roseanne (1997) *Adults count too: mathematics for empowerment*, Leicester, UK: National Institute of Adult Continuing Education (NIACE)

Bishop, A. (1988). Mathematical enculturation: a cultural perspective in mathematics education. Dordrecht, Holland: D. Reidel Publishing Co.

Buckingham, E. (1997). *Specific and generic numeracies of the workplace*. Melbourne: Deakin University.

Carnevale, A.P., Gainer, L.J., and Meltzer, A.S. (1990). *Workplace basics: The essential skills employers want*. San Francisco: JosseyBass.

Carraher, T.N., Schliemann, A.D., and Carraher, D.W. (1988). Mathematical concepts in everyday life. In G.B. Saxe and M. Gearhart (Eds.), *Children's Mathematics* (pp. 71-88). San Francisco: JosseyBass.

Coates et al. (1995). *National Reporting System*. Australia: Commonwealth of Australia and the Australian National Training Authority.

Coben, D. (2000). Numeracy, mathematics, and adult learning. In I. Gal (Ed.), *Adult numeracy development: Theory, research, practice* (pp. 33-50). Cresskill, NJ: Hampton Press.

Cockcroft, W.H. (1982). Report of the Committee of Inquiry into the Teaching of Mathematics in Schools. London: HMSO.

Cumming, J., Gal, I., and Ginsburg, L. (1998). *Assessing mathematical knowledge of adult learners: Are we looking at what counts?* (Technical Report TR98-05). Philadelphia, PA: University of Pennsylvania, National Center on Adult Literacy.

Curry, D., Schmitt, M.J., and Waldron, S. (1996, July). A framework for adult numeracy standards: The mathematical skills and abilities adults need to be equipped for the future. Boston, MA: The Adult Numeracy Practitioners Network.

Dossey, J.A. (1997). “Defining and measuring quantitative literacy”. In L.A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow’s America*. New York: College Entrance Examination Board.

European Commission. (1996). *White paper on education and training: Teaching and learning towards the learning society*. Luxembourg: Office for Official Publications of the European Commission.

Fey, James T. (1990). “Quantity” In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.

Forman, S.L. and Steen, L.A. (1999). *Beyond eight grade: Functional mathematics for life and work*. University of California, Berkeley: National Center for Research in Vocational Education.

Frankenstein, M. (1989). *Relearning mathematics: A different third ‘R’ – Radical maths*. London: Free Association Books.

Gal, I. (1993). *Issues and challenges in adult numeracy* (Technical Report TR93-15). Philadelphia, PA: National Center on Adult Literacy, University of Pennsylvania.

Gal, I. (1997). Numeracy: Imperatives of a forgotten goal. In L.A. Steen (Ed.), *Why numbers count: quantitative literacy for tomorrow’s America* (pp. 36-44). New York: The College Board.

Gal, I. (2000). The numeracy challenge. In I. Gal (Ed.), *Adult Numeracy Development: Theory, Research, Practice* (pp. 9-31). Cresskill, NJ: Hampton Press.

van Groenestijn, M. (1998). Constructive numeracy teaching as a gateway to independent learning. In Coben, D., and O’Donoghue, J. (Eds.), *Adults learning mathematics–4, Proceedings of the Fourth International Conference of Adults Learning Mathematics*. London: Goldsmiths University of London.

Heuvel-Panhuizen, M. van den, and Gravemeijer, K.P.E. (1991). Tests are not all that bad: An attempt to change the appearance of written tests in mathematics instruction at the primary school level. In L. Streefland (Ed.), *Realistic mathematics education in primary school*. The Netherlands: Utrecht.

Johnston, B. (1994, Summer). Critical numeracy? In *Fine print*, Vol.16, No. 4.

Jones, S. (1995). The distribution of literacy. In *Literacy, economy and society*. Ottawa: Statistics Canada.

Kirsch, I.S., Jungblut, A., and Mosenthal, P.B (1998). The measurement of adult Literacy. In S.T. Murray, I.S. Kirsch, and L.B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 105-134). Washington, DC: National Center for Education Statistics, US Department of Education.

Kindler, Jan, et al. (1996). *Certificates in General Education for Adults*. Victoria, Melbourne: Adult, Community and Further Education Board.

Lave, J. (1998). Cognition in practice: Mind, mathematics, and culture in everyday life. Cambridge, UK: Cambridge University Press.

Leonelli, E., Merson, M.W., Schmitt, M.J., and Schwendeman, R. (Eds.). (1994). *The Massachusetts ABE math standards project* (Vols. 1 and 2). Holyoke, MA: Holyoke Community College.

Marr, B., and Tout, D. (1997). A numeracy curriculum: Australian Association of Mathematics Teachers (AAMT) conference proceedings. Melbourne: AAMT.

Mathematics Functional Expert Group-PISA Study, Jan de Lange, Chair. (1998, June). *Framework for assessing mathematical literacy*. Draft technical report submitted to the OECD/Program for International Student Assessment (PISA) project.

McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In Grouws, D.A. (Ed.), Handbook of research on mathematics teaching and learning (pp. 575-596). New York: Macmillan.

Mullis, J.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Boston: Center for the Study of Testing, Evaluation, and Educational Policy.

Murnane, R., Willett, J., and Levy, F. (1995, March). *The growing importance of cognitive skills in wage determination* (Working Paper No. 50-76). Cambridge, MA: National Bureau of Economic Research.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.

Nunes, T. (1992). Ethnomathematics and everyday cognition. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 557-574). New York: Macmillan.

Oates, T. (1992). Developing and piloting the NCVQ core skill units: An outline of method and a summary of findings (Report 16). London: National Council on Vocational Qualifications (NCVQ).

Queensland Department of Education (1994). Unpublished manuscript

Resnick, L.B. (1987). Learning in school and out. *Educational Researcher*, *16*, 13-20.

Robitaille, D.F., and Travers, K.J. (1992). International studies of achievement in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 687-709). New York: Macmillan.

Rogoff, B., and Lave, J. (Eds.). (1984). *Everyday cognition: Its development in social context*. Cambridge, MA: Harvard University Press.

Rutherford, F.J and Ahlgren, A. (1990). *Science for all Americans*. New York: Oxford University Press.

Saxe, G.B. (1992). Culture and cognitive development: Studies in mathematical understanding. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saxe, G.B., Dawson, V., Fall, R., and Howard, S. (1996). Culture and children's mathematical thinking. In Sternberg, R, and Ben-Zeev, T. (Eds.), *The nature of mathematical thinking* (pp. 119-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schliemann, A.D., and Acioly, N.M. (1989). Mathematical knowledge developed at work: The contribution of practice versus the contribution of schooling. *Cognition and Instruction*, 6, 185-221.

Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools: A SCANS Report for America 2000*. Washington, D.C.: US Dept. of Labor.

Senechal, Majorie (1990) "Shape" In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.

Statistics Canada and The Organisation for Economic Cooperation and Development. (1996). *Literacy, economy, and society: First results from the International Adult Literacy Survey.* Ottawa: Author.

Statistics Canada, and Organisation for Economic Cooperation and Development. (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. Ottawa: Author.

Steen, L.A. (Ed). (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Research Council.

Steen, L.A. (Ed). (1997). *Why numbers count: quantitative literacy for tomorrow's America*. New York: The College Board.

Stein, S.G. (1995). *Equipped for the future: A customer-driven vision for adult literacy and lifelong learning.* Washington, DC: National Institute for Literacy.

Tobias, S. (1993). *Overcoming math anxiety*. New York: Norton.

Willis, S. (1990). *Being numerate: What counts?* Victoria: Australian Council for Educational Research (ACER).

Appendix 1

Sample items

The sample items presented on the following pages show examples of the tasks that were developed for the pilot study. These are tasks that have been discarded for the main survey; however, they provide an idea of the types of tasks used in the final selection of items.

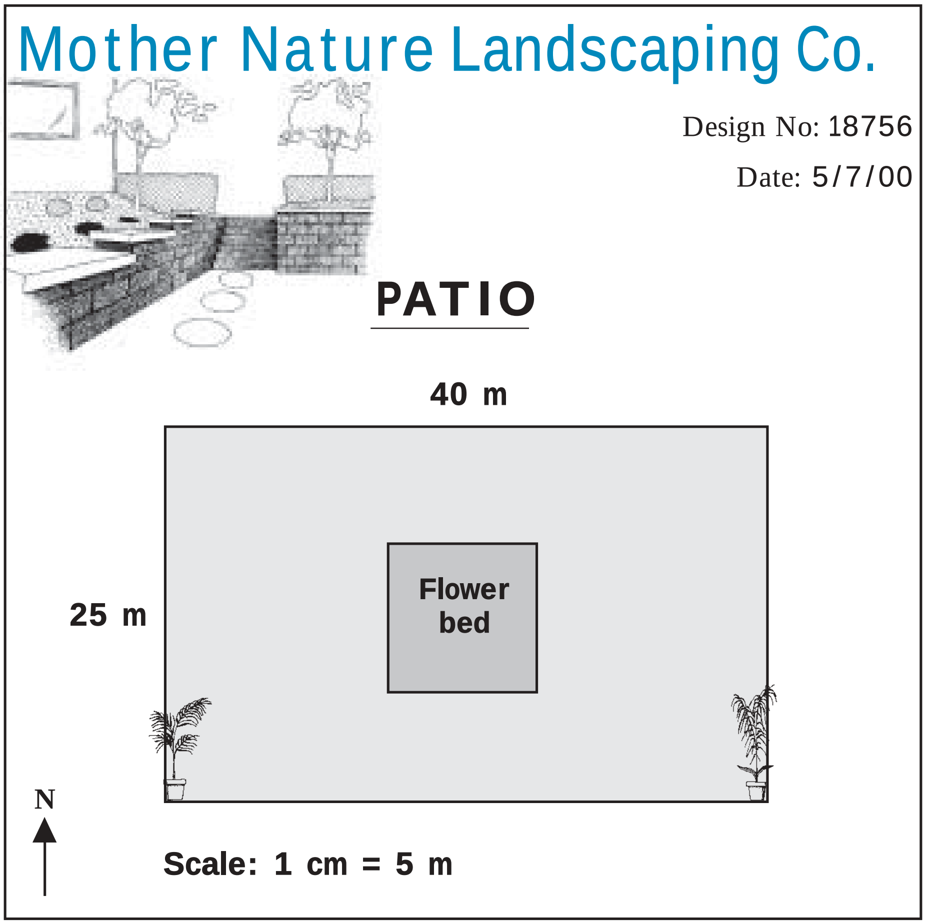
The tables accompanying them provide a broad analysis of each item that includes a few facet designations, the theoretical difficulty summary score that was assigned (ranging from 5, easiest, to 19, most difficult), and the difficulty statistics that were gathered during the pilot study.

Question

You wanted two pillows and bought them during the sale.

In percent, how much did you save?

Item name	Content	Response	Other feature	Theoretical difficulty	Percent correct
Pillows	change	interpret	picture, low text	12	77%



Questions

Assume that you work for Mother Nature Landscaping Co.

You are given the sketch showing a patio that is to be landscaped by your company.

The center square is to be planted with blooming flowers and the rest of the patio will contain a variety of potted plants arranged on its concrete surface.

a) The design calls for 5 large potted plants to be placed along the south edge of the patio, one in each corner (as shown) and 3 in between.

How far apart should the centers of the pots be if you want them to be equally spaced?

b) What is the area of the actual square flower bed in the middle of the patio? (Note: you can use the scale on the bottom).

Item name	Content	Response	Other feature	Theoretical difficulty	Percent correct
Landscaping-a	dimension, space	interpret, compute	diagram	15	32%
Landscaping-b	dimension	measure, compute, use formula	diagram with scale	15	38%

Question.

Which pack of film gives you more for your money?
Explain how you decided.

Item name	Content	Response	Other feature	Theoretical difficulty	Percent correct
Best buy	quantity	compute/ estimate explain	low text	7	66%

OPINION

LETTERS

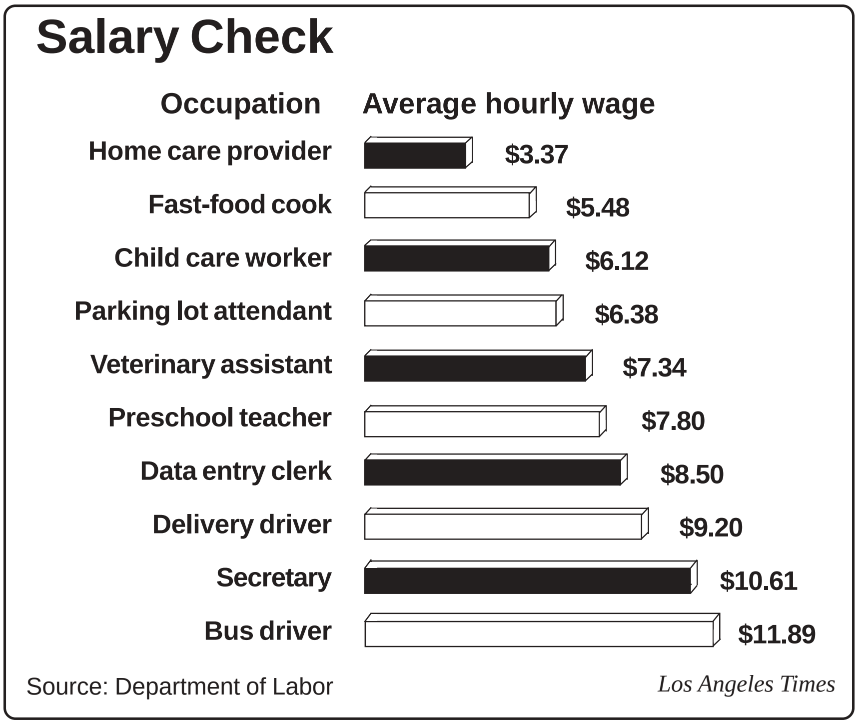
Last week we published the chart below on salaries. We received quite a few letters from readers. Some of the comments we include here.

From a secretary:
'I was interested to see your chart of average salaries. I was a homecare provider and changed jobs to a secretary last month.'

From an upset child care worker:
'The bus drivers have been on strike for 4 weeks for more pay. How dare they, the bus drivers already earn nearly double what child care workers earn per hour.'

And a final comment from a veterinary assistant who thinks we made an error:
'I am a veterinary assistant and make only $5.20 an hour. I think your chart must be wrong.'

Note from the editor: The source of data was the Department of Labor



Questions

Refer to the OPINION section of the newspaper.

a) What is the difference between the average hourly wage for a secretary and that of a home care provider?

b) How would you explain to the veterinary assistant that the average wage for veterinary assistants given in the chart can be correct?

Item name	Content	Response	Other feature	Theoretical difficulty	Percent correct
Wages-a	quantity, data	compute	locate in graph/text	10	78%
Wages-b	data	interpret, explain	Includes more text	15	27%

Appendix 2

Scoring for each of the complexity factors

Complexity Factor 1. Type of match/Problem transparency How difficult is it to identify and decide what action to take? How many literacy skills are required?		
Score 1	**Score 2**	**Score 3**
In the question and the stimulus, the information, activity or operation required: • is clearly apparent and explicit—and all required information is provided • is specified in little or no text, using familiar objects and/or photographs or other clear, simple visualizations • is about locating obvious information or relationships only • closed question—not open-ended	In the question and the stimulus, the information, activity or operation required: • is given using clear, simple sentences and/or visualizations where some translation or interpretation is required • is located within a number of sources within the text/activity. • fairly closed question	In the question and the stimulus, the information, activity or operation required: • is embedded in text where considerable translation or interpretation is required and/or • may need to be derived or estimated from a number of sources within or outside the text/activity and/or • the information or action required is not explicit or specified • more complex, open-ended task

Complexity Factor 2. Plausibility of distractors How many other pieces of mathematical information are present? Is all the necessary information there?		
Score 1	**Score 2**	**Score 3**
• no other mathematical information is present apart from that requested—no distractors	• there is some other mathematical information in the task that could be a distractor • the mathematical information given or requested can occur in more than one place • may need to bring to the problem simple information or knowledge from outside the problem.	• other irrelevant mathematical information appears • mathematical information given or requested appears in several places. • necessary information or knowledge is missing, so outside information or knowledge needs to be brought in

Complexity Factor 5. Expected number of operations How many steps and types of steps are required?		
Score 1	**Score 2**	**Score 3**
one operation, action or process	application of two or three steps, the same or similar operation, action or process	integration of several steps covering more than one different operation, action or process

Complexity Factor 3. Complexity of mathematical information/answer required

How complex is the mathematical information that needs to be manipulated?

Score 1	Score 2	Score 3	Score 4	Score 5
Context Based on very concrete, real life activities, familiar to most in daily life.	Based on common, real life activities.	Based on real life activities, but less often encountered.	Based on real life activities but unfamiliar to most	Based on abstract ideas or unfamiliar activity in a context new to most.
Quantity **Whole numbers** to 1,000 **Fractions, decimals, percents** • benchmark fractions ($^1/_2$, $^1/_4$, $^3/_4$) • decimal fraction for a half only (0.5) and equivalent as a percentage (50%)	• large whole numbers including millions • other benchmark fractions, like $^1/_3$ and $^1/_{10}$ • common decimals, like 0.1, 0.25 to 2 decimal places • common whole number percents, like 25% and 10%.	• large whole numbers including billions • other fractions • decimals to 3 decimal places (other than money) • all whole number percents	• negative integers • all remaining fractions, decimals and percentages	• all remaining types of rational (and some irrational) numbers including directed numbers
Pattern and relationship • very simple whole number relations and patterns	• simple whole number rates and ratios • whole number relations and patterns	• rates and ratios • relations and patterns including written everyday generalizations	• complex ratios, relations, patterns • simple formula	• formal mathematical information such as more complex formulae, knowledge of relationships between dimensions or variables, etc
Measures/Dimension/Space • standard monetary values • common everyday measures for length (whole units) • time (dates, hours, minutes) • simple, common 2D shapes • simple localised maps or plans (no scales)	• everyday standard measures length, weight, volume, including common fraction and decimal units • common 3D shapes and their representation via diagrams or photos • common types of maps or plans with visual scale indicators	• other everyday measures (area included) including fraction and decimal values • more complex 2D and 3D shapes, or a combination of 2 shapes • area and volume formulae • common types of maps or plans with ratio type scales	• all kinds of measurement scales • complex shapes or combinations of shapes	
Chance/Data • simple graphs, tables, charts with few parameters and whole number values • simple whole number data or statistical information in text	• graphs, tables, charts with common data including whole number percents—whole number scales in 1s, 2s, 5s or 10s • data or statistical information including whole number percents	• graphs, tables, charts with more complex data (not grouped data) • more complex data or statistical information including common average, chance and probability values • scales: more complex whole number, fractional or decimal	• complex graphs, tables or charts including grouped data • complex data or statistical information including probabilities, measures of central tendency and spread	

Complexity Factor 4. Complexity of Type of operation/skill How complex is the mathematical action that is required?				
Score 1	**Score 2**	**Score 3**	**Score 4**	**Score 5**
Communicate no explanation - a single simple response required (orally, or in writing)	• no explanation - a simple response required (orally, or in writing)	• simple explanation of a (level 1 or 2) mathematical process required (orally, or in writing)	• explanation of a (level 3) mathematical process required (orally, or in writing)	• complex, abstract and generative reasoning or explanation required
Compute • a simple arithmetical operation (+, -, x, ÷) with whole numbers or money	• calculating common fraction, decimal fraction and percentages of values • using common rates (e.g. $/lb.); time calculations; etc • changing between common equivalent fraction, decimal and percent values, including for measurements e.g. $^{1}/_{4}$ kg = 0.250kg	• more complex applications of the normal arithmetical operations such as calculating with fractions and more complex rates, ratios, decimals, percentages, or variables • simple probability calculations	• applications of other mathematical operations such as squares, square roots, etc	• more advanced mathematical techniques and skills e.g. trigonometry
Estimate	• estimating and rounding off (when requested) to whole number values or monetary units	• estimating and rounding off to requested number of decimal places	• making a contextual judgment re whether a found answer is realistic or not and changing the answer to the appropriate correct rounded (but not necessarily mathematically correct) answer.	
Use formula/model	• evaluating a given formula involving common operations (+, -, x, ÷)		• developing/creating and using straight forward formulae • using strategies such as working backwards or backtracking (e.g. 15% of ? = $255)	• generative reasoning • using and interpreting standard algebraic and graphical conventions and techniques
Measure • knowing common straight forward measures • naming, counting, comparing or sorting values or shapes	• visualizing and describing shapes, objects or geometric patterns or relationships • making and interpreting standard measurements using common measuring instruments	• using angle properties and symmetry to describe shapes or objects • estimating, making and interpreting measurements including interpolating values between gradations on scales • converting between standard measurement units within the same system	• calculating measures of central tendency and spread for non-grouped data • converting between non-standard measurement units within the same system • counting permutations or combinations	• converting between measurements across different systems
Interpret • locating/identifying data in texts, graphs and tables • orientating oneself to maps and directions such as right, left, etc	• reading and interpreting data from texts, graphs and tables • following or giving straight forward directions	• interpolating data on graphs • calculating distances from scales on maps	• generating, organising, graphing non-grouped data • extrapolating data • reading and interpreting trends and patterns in data on graphs, including slope/gradient	• graphing grouped data • calculating measures of central tendency and spread for grouped data

Chapter 6

ALL Problem Solving Framework

Jean-Paul Reeff *LIFE Research and Consult*
Anouk Zabal *LIFE Research and Consult*
Eckhard Klieme *German Institute for International Educational Research*

Table of contents

Table of contents

Foreword

This framework is based on the preliminary work of a European research network that worked on the development of new assessment tools for problem solving. Network partners were the Luxembourg Ministry for Education, the Center for School Development (Graz), the Max-Planck-Institute for Human Development (Berlin), the University of Groningen and the University of Heidelberg. Subsequent item development and accompanying work was carried out by a further network partner, the Institut für Bildungsforschung e.V. in Bonn (Germany).

Many colleagues contributed to this framework in different ways. The authors would like to express their gratitude to them all. Special thanks go to Judith Ebach and Joachim Funke for their valuable and substantial help towards improving this framework, and to Jules Peschar, Marieke van de Wal and Sietske Waslander for their early conceptual work.

1. Introduction

The Adult Literacy and Life-skills Survey (ALL) aims at assessing a broad range of skills important in everyday life and essential for social, professional and economic success. As we enter the 21st century, society changes at an increasingly fast pace and citizens are challenged by demands for greater flexibility as well as reflected and planned behavior. That problem-solving skills constitute one of the crucial life-skills is widely accepted. Problem solving is ranked as an important key qualification by labor market experts (see Binkley, Sternberg, Jones, and Nohara, 1999) as well as in the literature on vocational training and education (Didi, Fay, Kloft, and Vogt, 1993). Recent discussions of lifelong learning also point to problem solving as one of the major competencies to be fostered in a lifelong learning process. Furthermore, problem-solving skills were defined as an important outcome of schooling by OECD experts (OECD, 1997; Ryjchen and Salganik, 2000), and are often identified as high-level curricular aims (see, e.g., Svecnik, 1999). These few examples suffice to illustrate the importance of assessing problem-solving skills in the context of the ALL study.

Research on problem solving has a long tradition within both academic psychology and applied human resources research. One major challenge while developing a framework for problem solving for the ALL study was to adapt the findings of this research to the constraints imposed by a large-scale international comparative study. In order to do this, we will focus on an essential subset of problem solving, analytical problem solving.

The following chapters give a brief overview of problem solving from a scientific point of view and specify what is to be understood by analytical problem solving within the context of ALL. Once it is clearly established what is meant by analytical problem solving, the assessment rationale and assessment instruments for this domain are explained. Finally, initial results from the pilot study in five different countries will be presented together with a description of the problem-solving scale.

2. Problem solving from a scientific point of view

2.1 The concept of problem solving

2.1.1 Definition

For the purposes of the ALL study, we begin with a very general definition of problem solving as it is generally understood in psychological research (Hunt, 1994; Mayer, 1992; Mayer and Wittrock, 1996; Smith, 1991).

> *Problem solving is goal-directed thinking and action in situations for which no routine solution procedure is available.* The *problem solver* has a more or less welldefined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a *problem*. The understanding of the problem situation and its step-by-step transformation, based on planning and reasoning, constitute the process of *problem solving*.

This relatively broad definition makes it clear that problem solving is to be understood as a cognitive process. More specifically, we are exclusively concerned with cognitive-analytical problem solving. It is not to be confused with the intuitive everyday use of the term or with the clinical-psychological concept in which problem solving is associated with the resolution of social and emotional conflicts. Nevertheless, social context is also relevant for our definition of analytical problem solving, for example when problems have to be approached interactively and resolved through co-operation. Of course motivational factors such as interest in the topic and task-orientation also influence the problem-solving process. However, the quality of problem solving is primarily determined by the comprehension of the problem situation, the thinking processes used to approach the problem, and the appropriateness of the solution.

The *problem* itself can be characterized by different aspects:

- The *context* can reflect different domains, which may be of a theoretical or a practical nature, related to academic situations or to the real world. Within these domains, problems can be more or less authentic.
- The *scope* of a problem can range from working on limited, concrete parts of a task to planning and executing complex actions or evaluating multiple sequences of actions.
- The problem can have a well-defined or an ill-defined goal, it can have transparent (explicitly named) or intransparent constraints, and involve few independent elements or numerous interconnected ones. These features determine the *complexity* of the problem.

How familiar the context is to the target population, whether the problem involves concrete tasks or complex actions, how well the goal is defined, how transparent the constraints are, how many elements the problem solver has to take into account and how strongly they are interconnected – all these features will determine the level of problem-solving competency that is required to solve a certain problem. The empirical difficulty, i.e. the probability of giving a correct solution, will depend on the relation between these problem features on the one hand, and the subjects' competency level on the other hand.

The *cognitive processes* that are activated in the course of problem solving are diverse and complex, and they are likely to be organized in a non-linear manner. Among these processes, the following components may be identified:

1. Searching for information, and structuring and integrating it into a mental representation of the problem ("situational model").
2. Reasoning, based on the situational model.
3. Planning actions and other solution steps.
4. Executing and evaluating solution steps.
5. Continuous processing of external information and feedback.

Baxter and Glaser (1997) present a similar list of cognitive activities labeled "general components of competence in problem solving": problem representation, solution strategies, self-monitoring, and explanations.

Analytical problem solving in everyday contexts, as measured by the ALL problem-solving instrument, focuses on the components 1 to 3 (and to some extent 4).

2.1.2 Psychological models

Psychological models of these processes and the mental structures (representations) on which they operate have changed over the history of psychology, each tailored to the particular kinds of problems that were focused by the respective research paradigm. In the early years of cognitive psychology, for example, "insight" was seen as a major mechanism. This concept was appropriate in limited but ill-defined problem situations, where a sudden restructuring or reinterpretation of the problem yields the solution. Newell and Simon (1972), in their seminal book "Human problem solving", which served as a framework for numerous studies in cognitive information processing and artificial intelligence, described problem solving as a process of search in a "problem space" consisting of states (including given state and target state) and operators. This model was appropriate for the study of well-defined, "puzzle"-type problems. While Newell and Simon believed they had discovered rather universal mechanisms, research on scientific reasoning and expertise later proved that problem solving strongly depends on the use of domain-specific knowledge which was described in terms of rule systems, schemata, mental models or "mental tools" (see e.g. Chi, Glaser, and Farr, 1988; Weinert and Kluwe, 1987). At the same time, it became clear that meta-cognition plays a vital role for both problem-solving processes and the outcomes of problem-solving activities (Brown, 1987; Flavell, 1976). Meta-cognition is defined as the *process* of planning, monitoring, evaluating and regulating ongoing cognitions as well as the knowledge and beliefs *about* cognitive functioning.

In order to find out how well people can solve particular types of problems, it is not necessary to identify mental structures or process components in detail. Assessment frameworks need not meet the sophistication of cognitive-psychological models. However, even a purely functional approach to problem-solving assessment has to take into account some important results of psychological research, associated with the key terms "general intelligence", "complex problem solving", and "domain specificity". The following section provides a short overview of these findings and discusses the implications for the design of problem-solving assessments.

2.2 Problem solving, reasoning and related constructs

Problem solving as defined above is quite similar to some other constructs in modern psychology. Among them are "critical thinking" (Ennis, 1996; Norris, 1989), which encompasses judging the credibility of arguments, and naturalistic decision-making (Zsambok and Klein, 1997), defined as the use of knowledge and expertise to act under complex and uncertain conditions. Each of these constructs describes some kind of intellectual activity, based on reasoning and the application of knowledge. Therefore, they are closely linked to the construct of intelligence that many modern psychologists understand as a generalized capability to acquire, integrate, and apply new knowledge. Intelligence in turn is linked to more basic features of the human information processing system such as working memory capacity or mental speed (Neisser et al., 1996).

In the tradition of psychometric research, the core of general intelligence is called reasoning (Carroll, 1993) or information processing capacity (Süß, 1999). It is operationalized by tests using mathematical word problems, number series (e.g. 1, 2, 4, 7, 11, ... ?), and analogical reasoning, in particular by figural analogies like "/ is to \ as # is to ... ?". All these may be subsumed under the broad concept of problem solving as it was defined above – with the exception of rare cases in which highly trained persons solve such tasks using special algorithms. Thus, whatever indicator for problem-solving competence we use, it will to a certain degree be correlated to psychometric measures of reasoning ability. How strong this correlation is, and hence the extent to which problem-solving competence can actually be distinguished from reasoning, is an open question in cognitive-psychological research. Even with respect to complex, dynamic, computer-based problem-solving tasks (Frensch and Funke, 1995; see section 2.3 below) several studies suggest that inter-individual performance differences can be explained to a large extent by reasoning ability and basic features of the human information processing system (Süß, 1999).

In recent publications in the area of Differential Psychology, Robert Sternberg and his colleagues (see, e.g., Sternberg and Kaufman, 1998) have supported a very broad concept of intelligence, basically equating it with problem-solving abilities. Sternberg identifies three subcomponents of intelligence: a) analytical abilities such as "identifying the existence of a problem, defining the nature of a problem, setting up a strategy for solving the problem, and monitoring one's solution process", b) creative abilities "required to generate problem-solving options", and c) practical abilities needed to apply problem-solving strategies to real-life tasks. Sternberg assumes that practical intelligence is clearly discernible from analytical intelligence as assessed by means of the classic psychometric measures (IQ). However, methods for measuring creative problem-solving abilities and practical aspects of intelligence independently have yet to be devised. The procedure Sternberg proposed to measure practical intelligence cannot be regarded as a performance test: He presents respondents with descriptions of real-life or job-related problem situations and asks them to evaluate different response alternatives. If the evaluations made by the respondent correspond to those of a reference group ("experts" in an occupational field or representatively selected control groups for real-life problems), the respondent is said to have tacit knowledge, which Sternberg sees as the core of practical intelligence (Sternberg and Wagner, 1986).

The assessment of the third aspect in Sternberg's triarchic concept of intelligence – creativity – appears to be just as difficult. As problem solving involves new situations that cannot be dealt with routinely, it always requires a certain degree of creativity. Attempts to measure creativity independently as originality, flexibility and fluency of ideas (see Krampen, 1993) or to assess it as a distinctive feature of problem-solving performance (Mumford, Supinski, Baughman, Costanza, and Threlfall, 1997) have, however, yet to yield convincing results.

2.3 Addressing the complexity and dynamics of problem solving

In recent years, psychological research on problem solving has turned to increasingly complex, authentic problems with a broader scope (Sternberg and Frensch, 1991). It is no longer concerned with well-defined "puzzles" (in the extreme case, reasoning tasks as used in psychometric tests of human intelligence) that can be solved by applying suitable operations. Instead, it addresses the thinking of experts in scientific and professional domains (Reimann and Schult, 1996; Zsambok and Klein, 1997), planning and problem solving in reallife contexts (Funke and Fritz, 1995; Jeck, 1997; Lave, 1988), and the understanding and control of complex ecological, economic and technical systems (Dörner, Kreuzig, Reither, and Stäudel, 1983; Frensch and Funke, 1995). Computer simulation has proved to be an important tool for investigating complex problem-solving performance. In interaction with the computer, the problem solver explores the simulated system, generates and tests (more or less systematically) hypotheses about relationships and regularities, acquires knowledge, and may in this way learn to control the system by purposeful intervention. Systems used in computer-based research include realistic simulations of highly interconnected ecological or economic systems (Dörner, Kreuzig, Reither, and Stäudel, 1983), and systematically constructed, discrete, smaller-scale systems ("finite state automata"; Buchner and Funke, 1993) and virtual experimental environments. Adaptations of such systems in school contexts are described by Leutner (1992). From an educational psychology perspective, such procedures can be understood as environments for discovery learning (Boshuizen, van der Vleuten, Schmidt, and Machiels-Bongaerts, 1997; Leutner, 1992). From the perspective of problem-solving research, these instruments provide a new quality of problem tasks, distinguished by high levels of complexity and, in particular, by a dynamic character. These dynamic tasks have three advantages over static paper-and-pencil tasks:

1. The task *demands* are enhanced by an *active search and by the continuous processing of external information and feedback*. Although paper-and-pencil problem-solving tasks may also trigger the application, evaluation and – if necessary – the modification of processing strategies, the interaction with the computer makes such a course of action inevitable.
2. Computer simulations offer much more *authentic problem situations* than a written test.
3. Not only the results, but also the course of the problem-solving process can be recorded and assessed, i.e., the type, frequency, length and sequence of interventions made by the subjects. This provides *process-based indicators of problem-solving strategies*.

These three advantages demonstrate the benefits of using computers in the assessment of problem-solving performance. There are, however, serious theoretical and methodological problems when it comes to the measurement of strategies. The definition of such measures, their reliability, the extent to which they are comparable across different simulated systems, and the impact of motivational factors are research questions which are not yet adequately answered.

2.4 Domain-specific vs. general problem solving

One of the most important insights of recent research in cognitive psychology is that solving demanding problems requires at least some knowledge of the domain in question. The concept of a problem space through which a General Problem Solver moves by means of domain-independent search strategies (Newell and Simon, 1972) proved to be too simple to describe how problem situations are understood and the process of finding a solution. Efforts to identify a general, domain-independent competence for steering dynamic systems (operative intelligence) within the framework of complex problem-solving research were also unsuccessful; performance on such systems can only partially be transferred to other systems (Funke, 1991). However, research on grade-3 to grade-12 students showed that problem-solving skills clearly improve under well-tuned training conditions and that a substantial transfer across different problems can be achieved (Reeff et al. 1989, 1992, 1993; Regenwetter, 1992; Regenwetter and Müller, 1992; Stirner, 1993).

Problem solving is dependent on knowledge of concepts and facts (declarative knowledge) and knowledge of rules and strategies (procedural knowledge) in a given subject domain. Although it is obvious from past research that declarative knowledge in the problem domain can substantially contribute to successful problem-solving strategies, procedural knowledge is crucial as well. The amount of relevant previous knowledge available could also account for the relation between intelligence and problem-solving performance, as shown in the work of Raaheim (1988) and Leutner (1999). People with no relevant previous knowledge at all are unable to explore the problem situation or plan a solution in a systematic manner and are forced to rely on trial and error instead. Those who are already very familiar with the task are able to deal with it as a matter of routine. General intellectual ability, as measured by reasoning tasks, plays no role in either of these cases. When problem solvers are moderately familiar with the task, analytical reasoning strategies can be successfully implemented.

2.5 Analytical problem solving: Definition and levels of proficiency

Despite the appeal of computer-based assessment of problem-solving skills the operational constraints imposed by contemporary large-scale comparative studies necessitate focusing on aspects of analytical problem solving that can be measured by paper-and-pencil tasks. The present framework also refrains from addressing social, emotional, and creative aspects of problem solving.

> *Analytical problem solving* is the core of problem solving as a goal-directed cognitive process. It encompasses the use of content-specific and general knowledge, rules and strategies, and meta-cognition. A person's analytical problem-solving competency may be indicated by his or her performance in identifying a problem, searching for relevant information and integrating it into a coherent problem representation, evaluating the problem situation with respect to given goals and criteria, devising a plan – i.e. an ordered sequence of appropriate actions – and monitoring its execution.

Thus, analytical problem solving as it is defined here is closely related to reasoning ability and to the analytical subcomponent in Sternberg's triarchic theory of intelligence.

The approach described in this framework for the assessment of problem solving relies on the (moderately) familiar tasks mentioned in the previous section. Within a somewhat familiar context the problems to be solved are "intransparent" enough not to be perceived as pure routine tasks. On the other hand, the domain-specific knowledge prerequisites are sufficiently limited as to make analytical reasoning techniques the main cognitive tool for solving the problems.

As described in section 2.1, given a moderately familiar contextual embedding, the competency level required to solve a problem can be estimated by the scope and complexity of the problem. Based on results from general psychology and developmental psychology (e.g., Post-Piagetian theories of cognitive development), we can describe a hypothetical proficiency scale for analytical problem solving which was tested empirically with the ALL pilot data (see chapter 5). Similar levels have been described within the frameworks of other large-scale assessments of problem-solving competencies such as the project test for Hamburg/Germany (Ebach, Klieme and Hensgen, 2000) and the PISA 2003 assessment of cross-curricular problem solving (OECD, in press).

Four levels of problem-solving proficiency are postulated:

Level 1: At a very elementary level, concrete, limited tasks can be mastered by applying content-related, practical reasoning. At this level, people will use specific content-related schemata to solve problems.

Level 2: The second level requires at least rudimentary systematical reasoning. Problems at this level are characterized by well-defined, one-dimensional goals; they ask for the evaluation of certain alternatives with regard to transparent, explicitly stated constraints. At this level, people use concrete logical operations.

Level 3: At the third level of problem-solving proficiency, people will be able to use formal operations (e.g. ordering) to integrate multi-dimensional or ill-defined goals, and to cope with non-transparent or multiple dependent constraints.

Level 4: At the final and highest level of competency, people are capable of grasping a system of problem states and possible solutions as a whole. Thus, the consistency of certain criteria, the dependency among multiple sequences of actions and other "meta-features" of a problem situation may be considered systematically. Also, at this stage people are able to explain how and why they arrived at a certain solution. This level of problem-solving competency requires a kind of critical thinking and a certain amount of meta-cognition.

The next chapter gives a brief overview of different tasks and approaches for measuring problem-solving competency. Based on the discussion in chapter 3, a concrete solution for measuring *analytical problem solving*, as it was just introduced, is presented in chapter 4.

3. Measurement of problem solving

There are at least three different sources for the design of problem-solving tests: Tasks used in psychological research, domain-specific problem-solving tasks, and tasks used in previous large-scale assessments of cross-curricular or practical problem solving. These three possibilities will be examined in the following sections.

For the ALL study, the following requirements should be examined: The extent to which tasks tap broad analytical problem-solving abilities and are in this respect theoretically sound. They should furthermore be embedded within a real-life context that is realistic enough to trigger actual and not artificial problem-solving processes, and a context that does not make specialized knowledge a prerequisite. Finally, they should show adequate psychometric properties and be compatible with the constraints imposed by a large-scale assessment.

3.1 Tasks used in psychological research on problem solving

During the 20th century, psychological research in the area of problem solving concentrated on a few experimental paradigms. For example, the famous radiation problem in cancer therapy (Duncker, 1945), the water-jug problems (Luchins, 1942), the "Tower of Hanoi" (Newell and Simon, 1972) and it's analogies, Wason's rule induction task (Wason, 1966), traveling-salesman problems or Cryptarithmetics were used again and again in experimental settings (cf. Anderson, 1999). In addition to these puzzle-like problems, psychologists used knowledge-rich tasks such as chess games, geometry problems, algebraic word problems, mechanical reasoning or computer programming. In the European tradition of problem-solving research, computer simulations of various economical or ecological scenarios were introduced as a means of investigating human behavior in ill-defined, dynamic, intransparent, and complex problem situations (see Frensch and Funke, 1995). Thus, one possible strategy for the design of problem-solving assessment instruments could be to implement one or more of these paradigms. However, the tasks used in experimental research a) are often well-known to a larger public, b) are not appropriate for large-scale assessment, c) are not tailored to the life and experiences of the target population. Thus, the challenge would be to adapt these tasks, transform them into appropriate test formats, and contextualize them in a way that is meaningful to subjects across participating countries. The heterogeneity of the tasks brings up another problem: Mixing for example a Tower-of-Hanoi-like problem with an "insight"-problem would most probably yield a test with low internal consistency and unknown validity.

3.2 Domain-specific problem-solving tasks

Problem solving can occur in any domain, and there are abundant domain-specific problem-solving tests, especially in the areas of educational and vocational research. The most interesting tests are those that use innovative formats such as the "Clinical Reasoning Test" (Boshuizen et al., 1997), which is based upon case studies in patient management, the "Overall-Test" of complex, authentic decision-making in business education (Segers 1997), or the "What if Test" which measures intuitive knowledge acquired in exploratory simulations of science phenomena (Swaak and de Jong, 1996). For science, Baxter and Glaser (1997), provide a systematic approach to performance assessment tasks, allowing for an analysis of cognitive complexity and problem-solving demands. Within the domain of mathematics, there is a long tradition of research on problem-oriented thinking and learning (Hiebert et al., 1996; Schoenfeld, 1992), and related assessment strategies (Charles, Lester and O'Daffer, 1987; for an integrated discussion from an educational, cognitive-psychological and measurement perspective see Klieme, 1989). Collis, Romberg, and Jurdak (1986), for example, developed a "Mathematical Problem-Solving Test" which used so-called "super-items", each composed of a sequence of questions that address increasing levels of cognitive complexity. Since the seminal work by Bloom and colleagues (Bloom, Hastings, and Madaus, 1971), there have been various attempts at differentiating task complexity levels, a more recent example being the SOLO taxonomy (Collis et al., 1986). It is interesting to note that the former literature on taxonomy of learning objectives and related tasks did not contain any category like "problem solving", because Bloom and his colleagues conceptualized problem solving as an integration of all the levels they proposed (reproduction, understanding, application, and so on).

3.3 Tasks used in previous large-scale assessments

Recently, several attempts have been made to implement measures of cross-curricular problem solving in large-scale assessments. A general test of cross-curricular competence developed by Meijer and Elshout-Mohr (1999) is based on critical thinking inventories. This test is promising but measures quite heterogeneous concepts. Sternberg's practical cognition test and a problem-solving test developed by Baker (1998), O'Neil (1999) and colleagues at the Center for Research on Evaluation, Standards and Student Testing (CRESST) in the United States have also been piloted in large-scale assessments. In the pilot version of Sternberg's test, the respondents are provided with a list of workplace or everyday-life related problem situations, and asked to choose one of several possible solutions. The respondent's answer pattern is then compared to the average pattern in their culture. The degree of agreement between the respondent's choice and the national representative sample is regarded as an indicator of the respondent's "common sense". This means that there is no "right" or "wrong" solution to the problems. The CRESST problem-solving test is based on a framework that defines domain-dependent strategies, meta-cognition, content understanding, and motivation as components of problem solving. Note that only the first two of these are understood as aspects of problem-solving competence in the present framework, whereas the third and fourth are regarded as *prerequisites* that need to be measured independently. To assess strategies, the CRESST authors confronted respondents with information on a technical device (tire pump) or a similar biological system, which was described as malfunctioning, and asked them to think about trouble-shooting actions. To assess content understanding subjects were asked to explain how the device works by drawing a "knowledge map". Several field trials showed that in principle the instrument is feasible, although its difficulty (with less than 25% of the adults being able to solve the trouble-shooting problem) and reliability were not sufficiently convincing.

Trier and Peschar, as a part of their work for OECD-Network A (OECD, 1997), analyzed problem solving as one of the important cross-curricular competencies. They constructed an item to measure skills in written communication. The respondent is asked to plan a trip for a youth club. This essay-like planning task is based on "in-basket" documents. The task proved to be too difficult for the target population, and low levels of objectivity were reached in scoring the answers.

The idea of using planning tasks to measure problem solving was independently invented by other research groups. Funke and Fritz (1995) devised several experimental variants of tests requiring planning skills. Klieme, Ebach et al. (in press) developed a multiple-choice test of problem-solving competencies for a large-scale assessment program in one of the German federal states. This test consists of "projects" such as *organizing a party* or *planning a trip* that involve different types of tasks. This "project approach" is currently also used to assess cross-curricular problem-solving competencies in a German extension to PISA (Klieme, Funke et al., 2001; Klieme, in press). On an international scale, similar approaches have been recently used within the OECD's framework for problem-solving assessment (OECD, in press). PISA 2003 will include so-called "decision making" problems as well as "system analysis and design" problems. Both problem-solving types share the main features of "projects" as defined here.

3.4 Conclusions

This overview of the assessment instruments available to measure problem-solving competencies shows that none of these instruments fulfill all of the requirements for the problem-solving assessment within the ALL study. However, the project approach retains the potential to be a powerful instrument for assessing analytical problem-solving skills in real world context, as will be elaborated in greater detail in the next chapter. Solving problems in project-like settings is important and relevant for adults in both their professional and their private life. Furthermore, the project approach has been successfully implemented in other large-scale assessments, and it can be realized as a paper-and-pencil-instrument, which is of crucial importance for contemporary large-scale surveys. The item construction allows for a distinctive operationalization of the various proficiency levels defined in section 2.5. As described in chapter 5, the results of the feasibility study and the pilot study show the success of this approach.

The next chapter gives an in-depth description of the project approach as implemented for the ALL study.

4. The "Project approach"

How can contextualized, real-life problems be defined and transformed into test items? The project approach put forth in this framework uses different problem-solving phases as a dimension along which to generate the actual test items.

Following Pólya (1945, 1980), the process of problem solving has been frequently described in terms of the following stages:

1. Define the goal.
2. Analyze the given situation and construct a mental representation.
3. Devise a strategy and plan the steps to be taken.
4. Execute the plan, including control and – if necessary – modification of the strategy.
5. Evaluate the result.

These stages correspond to the results of research on vocational training and job analyses within educational research and applied psychology that have been described as a part of the so-called "complete action" approach. Extensive analyses of very different jobs (different professions with varying types of work places) indicate that new forms of labor organization require people to perform more complex operations that go "beyond mere routine". Nowadays, even production workers and office clerks are required to master complex tasks requiring integrative skills. Complete actions include different steps such as planning, executing and evaluating. The basic structure of the model of complete action is thus fully compatible with the above-mentioned normative process model for problem solving – action steps are similar to problem-solving steps.

The model of complete action has been successfully applied to curriculum development, assessment, and certification reforms in various professions in both Germany and Luxembourg (Hensgen and Blum, 1998; Hensgen and Klieme, 1998). The main idea is that both training tasks and also test problems should include all or most elements of a complete action.

4.1 ALL problem solving test construction

The ALL project approach uses this complete action model to establish the underlying structure of the problem-solving test. The different action steps define the course of action for an "everyday" project. One or more tasks or items correspond to each of these action steps. The respondents thus work on the individual tasks that have been identified as steps that need to be carried out as a part of their project (e.g., "planning a family reunion" or "renovating a clubhouse").

Embedding the individual tasks in an action context yields a high degree of context authenticity. A project, designed as a complete action, encompasses various tasks that can vary in complexity. Since assessing problem-solving skills is a relatively new endeavor, it seems appropriate to provide a somewhat detailed account of the general construction process. Furthermore, such a description is important for transferring or adapting this concept for other assessment purposes. Figure 1 illustrates the basic construction of a project.

Figure 1

The project

The Project						
Define the goal - analyze the situation - plan the solution - execute the plan - evaluate the result						
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7

Although they are all part of a comprehensive and coherent project, the individual tasks are designed so that they can be solved independently of one another. This is necessary in order to fulfill the requirements of the underlying measurement model.

Table 1 provides an overview of the problem-solving steps corresponding to the above-illustrated action steps. Different components and aspects of each of the problem-solving steps are listed.

Table 1

Problem-solving steps

Define the Goals	• Set goals. • Recognize which goals are to be reached and specify the essential reasons for the decision. • Recognize which goals/wishes are contradictory and which are compatible. • Assign priorities to goals/wishes.
Analyze the Situation	• Select, obtain and evaluate information. → What information is required, what is already available, what is still missing, and what is superfluous? → Where and how can you obtain the information? → How should you interpret the information? • Identify the people (e.g. with what knowledge and skills) who are to be involved in solving the problem. • Select the tools to be used. • Recognize conditions (e.g. time restrictions) that need to be taken into account.
Plan the Solution	• Recognize which steps need to be taken. • Decide on the sequence of steps (e.g. items on the agenda). • Coordinate work and deadlines. • Make a comparative analysis of alternative plans (recognize which plan is suitable for reaching the goals). • Adapt the plan to changed conditions. • Opt for a plan.
Execute the Plan	• Carry out the individual steps (e.g., write a letter, fill in a form, make calculations).
Evaluate the Results	• Assess whether and to what extent the target has been reached. • Recognize mistakes. • Identify reasons for mistakes. • Assess consequences of mistakes.

The construction of the problem-solving assessment instrument involved developing a pool of projects with different real-life contexts and a variety of item formats. Experts from several countries provided input for the development of different projects. Based on this, eight projects were initially developed and the best four were selected after a feasibility study and modified for the final instrument used in the pilot study. The different phases that directed the development of the problem-solving instrument and some of the criteria that are relevant to this process are illustrated in figure 2.

Figure 2

Five phases of test construction

Phase 1 Selecting the Subject for the Project

Chose a problem that

- is suitable for adults ages 16 to 65 and with varying educational backgrounds
- is relevant to everyday life for the greatest possible number of persons within the target group, and can generate an adequate problem context
- requires a level of general knowledge that an entry-level worker would typically have (and not more)
- is domain-independent
- is politically acceptable
- can be transferred to other cultures
- has no group-related bias (e.g. no gender bias)
- allows at least six mutually independent items to be created
- covers most of steps in the process model

Phase 2 Describing a Problem Situation and Sketching out the Action Steps

After drafting the problem situation, sketch out a sequence of actions that could form a part of solving the problem. This sequence of actions should approximately follow the five steps of the process model and yield a realistic sequence that would make sense in everyday life.

Phase 3 Developing Items Consistent With the Action Steps

- Develop a rudimentary task structure
- Aim at tapping certain processes and a certain level of difficulty
- Formulate the general problem and question
- Select adequate information to be processed (relevant, distracting, and irrelevant but contributing to the coherence of the problem context)
- Select an item format
- Develop response alternatives (correct answers and distractors)
- Create a solution key and scoring guides

Phase 4 Examining and Revising the Project

The projects and tasks are then revised and modified by experts according to the following criteria:

- General logic of action steps
- Coverage of relevant cognitive processes
- Completeness of the information provided
- Clarity in the statement of the item
- Unequivocal solutions
- Balance among the item formats
- Independence of the questions from each other
- Life relevance/face validity
- Degree of difficulty appropriate to the target group
- Amount and type of previous knowledge required
- The appropriateness in other cultures
- Political correctness
- Lack of group-related bias

Phase 5 Feasibility Study

A feasibility study with small samples should provide first results on:

- Consistency, reliability, difficulty, and discriminative power of the individual items
- Internal structure of the test
- Adequacy with respect to gender, education level, socio-economic status, motivation, previous experience, and language/culture
- Acceptance of instrument
- Operational concerns

Based on these results, the projects should then be revised and modified.

The thus constructed projects are structured in the following way:

1. A description of the problem situation is given, i.e. the “project” is introduced and the respondent’s role is specified, followed by a list of steps that need to be worked through as a part of this “project”. This list of things that need to be done corresponds exactly with the items the respondent is then asked to solve.
2. The different tasks: The information section supplies an introductory description and provides all the information the respondent needs to solve the task. The question section specifies the concrete question or questions (there can also be multiple, combined questions) that the respondent is requested to answer, and specifies how these are to be answered.

According to these development guidelines eight projects were designed for potential use in a large-scale survey. A feasibility study in Germany, with samples of approximately 60 respondents per tested project, yielded first satisfactory results. Theoretical considerations and the analysis of the empirical results led to some modifications of the projects and their tasks. A second feasibility study in the United States, with samples ranging from approximately 125 to 210 respondents per tested project, provided more ample insight into how well the projects worked empirically. Together with the feedback from the respondents, the analysis of the results of this second pre-pilot study shaped the way for further developing the projects and making a final selection and modification of the projects to be used in the pilot study.

4.2 Example

The following example illustrates a concrete realization of a project. For this purpose a project that is not included in the final ALL instrument is introduced and one typical problem-solving task is shown. The project is about "Planning a trip and a family reunion".

In the introductory part of the project, the respondent is given the following summary describing the scenario and overall problem:

> *"Imagine that you live in City A. Your relatives are scattered throughout the country and you would like to organize a family reunion. The reunion will last 1 day. You decide to meet in City B, which is centrally located and accessible to all. Since you and your relatives love hiking, you decide to plan a long hike in a state park close to City B. You have agreed to be responsible for most of the organization."*

The respondent is then given a list of steps he or she needs to work through, in this example the following list:

- *Set the date for the reunion*
- *Consider your relatives' suggestions for the hike*
- *Plan what needs to be done before booking your flight*
- *Answer your relative's questions about traveling by plane*
- *Book your flight*
- *Make sure your ticket is correct*
- *Plan the trip from City B to the airport*

The first task of this project "Set the date for the reunion" is a good example of a typical problem-solving task and is now shown as it would appear in a test booklet.

Example task: Set the date for the reunion

! ***The family reunion should take place sometime in July.***

You asked all your relatives to tell you which dates would be suitable. After talking to them, you made a list of your relatives' appointments during the month of July. Your own appointment calendar is lying in front of you. You realize that some of your relatives will have to arrive a day early in order to attend the family reunion and will also only be able to return home on the day after the meeting.

Please look at the list of your relatives' appointments and your own appointment calendar.

List of your relatives' appointments in July 1999

Henry	Karen	Peter	Janet	Anne	Frank
Vacation in City E beginning on July 26; Appointment on July 11	Every day of the week is okay except Thursdays and on July 16	Business appointments on July 2, July 13, and between July 27 and 29	Doesn't have any appointments	Unable to attend reunion on July 5, July 20, or July 24	Has to be away sometime during the 1st full week in July on business, but will find out the exact dates shortly before

Henry, Karen, and Peter could arrive on the same day as the reunion whereas Janet, Anne, and Frank can only arrive on the afternoon before and return home on the day after the reunion.

Example task (cont.)

Your appointment calendar for July 1999

July 1999

Thurs.	1	*Meeting with David*
Fri.	2	
Sat.	3	
Sun.	4	
Mon.	5	
Tue.	6	
Wed.	7	
Thurs.	8	
Fri.	9	
Sat.	10	*Hike in City C*
Sun.	11	
Mon.	12	
Tue.	13	
Wed.	14	
Thurs.	15	
Fri.	16	
Sat	17	
Sun.	18	
Mon.	19	
Tue.	20	
Wed.	21	
Thurs	22	
Fri.	23	
Sat.	24	
Sun.	25	
Mon.	26	
Tue.	27	
Wed.	28	*Vacation*
Thurs	29	*Vacation*
Fri.	30	*Vacation*
Sat.	31	

Question 1. Which of the following dates are possible for the family reunion?

Please select all possible dates.

a	July 4
b	July 7
c	July 14
d	July 18
e	July 25
f	July 29

Firstly, this project illustrates nicely how the action steps logic is actually "translated" into a concrete thematic action flow. The underlying plot – planning a trip and a family reunion – constitutes a very typical everyday-type of action that presumably a large majority of people in different countries will be able to relate to. The action steps themselves and their sequence can deviate from the normative complete action model, as is the case here. The normative model is used as a guideline that is adapted to each specific context. In this case, for example, the task "Consider your relatives´ suggestions for the hike" corresponds approximately to the action step "analyze the situation", the task "Plan what needs to be done before booking your flight" corresponds to the action step "plan the solution", and "Book your flight" is a typical example for the action step "execute the plan".

The example task gives a first indication of item structures and formats. The tasks typically start off with a short introduction to the situation, followed by varying types and amounts of information that need to be worked through. In the example task, in order to set the date for the family reunion, the respondent needs to process, compare and integrate the information provided in the list of the relatives' appointments, including the addendum to this list, and their own appointment calendar. Here the information is mostly textual and in form of tables. The answer format is a multiple-choice format with more than one correct response alternatives, although the number of correct response alternative is not specified.

4.3 General item description and response formats

This section gives an overview of some of the item formats and other item elements found within all the projects. To recapitulate, the aim of the problem-solving development team was to design a test that measures the cognitive processes essential to analytical problem solving and has high acceptance and appeal for a large number of adults in different cultures. The items need to tap the pertinent cognitive processes in a rich variety of ways and with a well-balanced range of difficulties. In order to do this, a broad range of task types with a large thematic variety and different task demands was developed. These tasks involve for example identifying relevant information, interpreting this information in terms of the task demands, using practical reasoning, relating different elements and judging the correct correspondence of different elements. More difficult variants could involve analytically analyzing the constraints and interconnections between different relevant elements and for example ordering the information elements according to the given constraints, generating combinational solutions to given problems, etc.

As previously mentioned, one criterion for the item construction was that item formats in the resulting item set are well-balanced. In order to measure the different relevant cognitive processes in a diversified manner, the items should ideally involve different amounts and types of information (textual, graphical, numerical, etc.), and different response formats. The response formats implemented may be summarized as follows:

- Multiple-choice items

  Different response alternatives are given, and the respondent is asked to mark the correct response alternative(s). The simplest multiple-choice items require the respondent to mark one response alternative. Some multiple choice items have more than one correct response alternative, and the number of correct responses may or may not be specified.

  The response alternatives themselves also have varying formats: They may appear as statements, numbers, drawings, sequences of steps, and so on.

- Non-standard closed-format items

  These kinds of items usually present some kind of a table in the question section, and the respondents´ task is to mark one or several of the cells in the table. The respondent has to associate the two given dimensions and select the correct combinations. An example of a typical question would be deciding which persons should carry out particular chores.

- Open-answer items

  The respondents are required to generate their own answer and write it down in the space provided. This may involve writing down one or more numbers or letters or combinations of both, filling in forms, or specifying the required information. For example, the respondents may have to write down a price or a date or specify errors they found. They may be asked to indicate the sequence in which they would carry out certain actions. Sometimes respondents have to give an explanation for a response they gave previously.

The majority of the items have either a multiple-choice or some other non-standard closed format. The nature of the tasks and the relative intransparency of the problems posed are linked with a generally relatively high difficulty level. However, this can be somewhat counteracted by structuring the answer possibilities in this manner. It should also be noted that in doing so the data-processing load is greatly reduced and is also less error prone. Including a restricted number of open-format items ensures a higher life-relevance and broadens the scope of the test.

4.4 Conclusions

As presented above, the project approach for the conceptualization of essential subsets of problem-solving competency particularly aims at analytical problem solving in well-defined, contextualized problem situations. The model of problem solving as outlined above serves as a framework for item development and puts analytical problem-solving tasks in context.

Solving the project tasks requires analytical operations such as searching, understanding, systemizing, organizing, evaluating, reasoning and combining information. These cognitive operations are essential for problem solving, defined as an information processing activity. Additionally, the tasks sometimes demand a certain kind of practical reasoning, which can best be described as the application of common sense or everyday knowledge.

Of course, there are some aspects of problem solving that cannot be measured within this approach: The dynamic aspects of task regulation (continuous processing of incoming information, coping with processes that cannot be influenced directly, coping with feedback and critical incidents) can only be addressed by computer-simulated tasks (Complex Problem Solving). The motivational, affective, and self-regulatory aspects of task regulation although implicit in the test tasks can only be explicitly addressed in a questionnaire or some similar method.

Problem solving behavior triggered by this test will depend on general, context-specific, domain-specific, and situation-specific processes. Nevertheless, this test is designed to tap a general (latent) competency for analytical problem solving as an essential part of problem solving. In fact, such a latent dimension has been established in large-scale assessments among student populations that used the project approach (Klieme, Ebach et al., in press; Klieme, Funke et al., 2001), and the data of the ALL pilot study corroborate these results, as will be reported in the next chapter.

5. Defining the ALL proficiency scale for problem solving: Empirical results from the field trial

5.1 Design of the ALL pilot study, and scaling and validity issues

Based on the development work outlined in the previous chapter and the subsequent results of a feasibility study in Germany and the United States, four projects were selected for an extensive pilot study in a number of countries. For each of these projects, a long and a short version were created in order to provide for more flexibility and to check if a shorter and more economical version of the projects yields data of equivalent quality and comparable measures for problem-solving competencies.

Thus four projects, each with a long and a short version, were tested in the ALL international large-scale pilot study. Two of the projects revolve around planning some kind of public event. The respondent is instructed to imagine she is a volunteer in a local group who has decided to organize a major event. The respondent is then led through a sequence of 3 to 7 tasks that correspond to action steps. The exact number of tasks or steps depends on the project and the version (short vs. long) she is working on. As explained in the last chapter, these action steps break up the general aim of the project (in this case "organizing an event") into setting a goal, analyzing, planning, executing, and evaluating. Within each of the action steps, the respondent is provided with some pieces of information – e.g. program of the event, time tables, places, people involved, budgets, and so on. The respondent needs to put a given set of actions into the right order, to draw conclusions and to make decisions based on the information given. In a similar manner, the two other projects implemented in the pilot study address complex decision making processes in a private context, e.g. making an important purchase.

The pilot study took place in five countries and in six languages: Belgium (Flemish), Canada (French and English), Italy, Norway and Switzerland (French, German, and Italian). The preliminary analyses discussed here are based on data from four countries (excluding Belgium), covering seven test versions in five languages. The total number of respondents was N=2102.

A problem-solving test booklet consisted either of two 30-minute problem-solving projects (long versions), or of four 15-minute projects (short versions). Therefore, respondents worked either on two long versions of the projects, or all four short ones. The long versions were labeled E, F, G and H, and the respective short versions were labeled I, J, K and L. Approximately equal numbers of respondents worked on each of the combinations E + F, F + G, G + H, and H + E (so-called matrix design), while twice this number of respondents worked on I + J + K + L. Because of design issues, missing values and some deviations from the standard implementation, case numbers varied from analysis to analysis. The preliminary analyses reported here are based on data sets of 530-710 respondents per item with a mean of 668 respondents per item.

The general aims of this pilot study were to:

1. test whether there is a common scale across projects.
2. evaluate empirical and other arguments pro and contra the short and long versions of the problem-solving projects, and make a recommendation for one of the versions.

3. select and revise items for the main survey.
4. validate the proficiency scale, especially the four levels of problem-solving competency that were introduced in chapter 2.

Methods from classical test theory as well as advanced models from item response theory (IRT) were applied. In addition, experts classified the items and rated item features. Most of the analyses were done in preparation of the final item selection and revision. Therefore, some of the results presented below are based on preliminary data sets and preliminary scoring procedures. However, comparisons carried out after the final test analysis revealed that the results are stable. For example, the correlation between our first version of item difficulty parameters and the final parameters, generated at Educational Testing Service (ETS) after the item selection and scoring had been finalized, is .92.

The results of the pilot test are quite conclusive. In the following, the relevant criteria and analytical procedures for each of the four issues will be described, followed by a short presentation of the corresponding findings from the ALL pilot study. Thus, it will become clear how the pilot results were used to select a final set of instruments and to develop an optimal instrument for the assessment of analytical problem solving.

5.2 A unique, common scale for analytical problem solving

5.2.1 Criteria and expectations

The matrix design of the field trial allowed for an integrated analysis of the long versions of all four projects. Thus, it was possible to estimate the latent (error-free) correlations between the projects. We expected these latent correlations to be around or above .90. Correlations of this size would indicate that the four projects could, in fact, be interpreted as building blocks of a single, common latent dimension.

The classical approach to test and item analysis – calculating item-test-correlations and estimating test reliability by the so-called coefficient alpha – could be applied to the combined short versions of all four projects (I+J+K+L), as these 18 items were administered to the same group of respondents. According to standards of item construction, an alpha coefficient above .80 would indicate that all the items – regardless of the different project contexts – make up a single, consistent dimension.

5.2.2 Findings and conclusions

The calculated pair-wise latent correlations between the different blocks ranged from .925 to .959. The combined short versions show a sufficiently high consistency (alpha = .81; part-whole-corrected item–test-correlations from .23 to .55 with a median of .38).

Thus, we can conclude that the items from all four projects form a common latent dimension, i.e. the analytical problem-solving scale. This is true both for the long and short versions of the problem-solving instrument This finding is very much in line with results from earlier implementations of the project approach (Ebach, Klieme and Hensgen, 2000, Klieme et al., 2001), where structural equation models (SEM) showed that problem-solving tests based on the project approach make up a unique dimension. This result has important consequences for the validity of the ALL problem-solving scale. It shows that the problem-solving test does not merely measure the ability to cope with certain special, context-dependent planning problems. Instead, the items actually do tap a general competency for analytical reasoning and decision making in complex situations where problem solving is required.

5.3 Long vs. short version of projects

5.3.1 Criteria and expectations

From an operational and financial point of view, the short versions of the projects were considered more desirable than the long versions. Therefore the critical question was whether the short versions are as informative as the long versions:

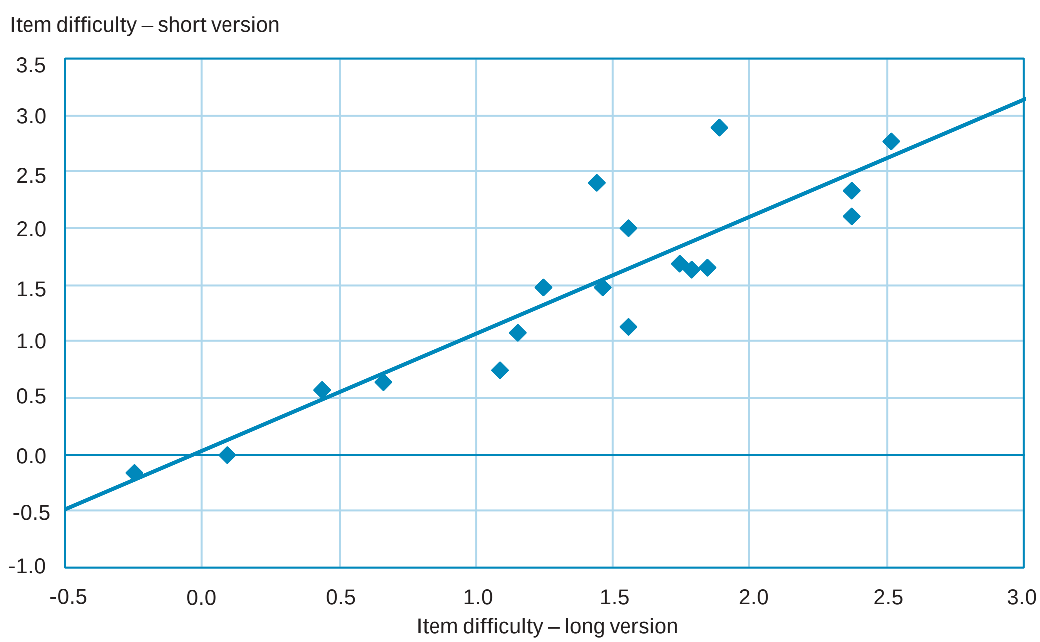
- Do they measure the same dimension, i.e. can the item parameters of the short versions be linearly transformed into the parameters of the long versions?
- Is the reliability of the combined short versions comparable to that of the long versions?
- Do the combined short versions cover a similar range of difficulties and proficiency levels?

5.3.2 Findings and conclusions

The following diagram visualizes the item difficulty[1] parameters of the 18 items that were included both in the short and in the long versions of the problem-solving projects. All points in the diagram (each representing a single item) are close to a straight line, indicating that there is an almost perfect linear dependency between the difficulty estimated from the short version and the corresponding difficulty index estimated from the long version. The correlation across 18 items is .90. Thus, both versions represent the same underlying construct.

Figure 3

Comparison of item difficulties for long and short versions



1. Within ALL, the so-called P80 index is used to operationalize item difficulty. It indicates the point on the latent competency scale (with logits as units of measurement) where the probability of solving the item correctly equals 80 percent.

With a coefficient alpha of .81, the test consistency for the integrated short versions exceeded all the consistency indices for the long versions. The best alpha coefficient found in the pilot study for the long version of one single project was .76. However, the individual long versions of the single projects had a smaller number of items than the four short versions together. Estimating the alpha that would have resulted from a single 20-item project leads to a coefficient of .90. Thus, the integration of short versions of several projects results in a somewhat lower consistency than a test that is made up of the same number of items from one single project. However, the difference is not that large, and .81 is normally thought to be an acceptable alpha value.

From these findings it was concluded that the problem-solving test for the main survey should be based on the combined short versions of all projects. With regard to the range of item difficulties, the short versions showed a somewhat more restricted range compared to the long versions. Therefore, it was decided to add some of the items from the long versions to the short-version booklet for the final item selection, as described in the next section.

5.4 The final item selection

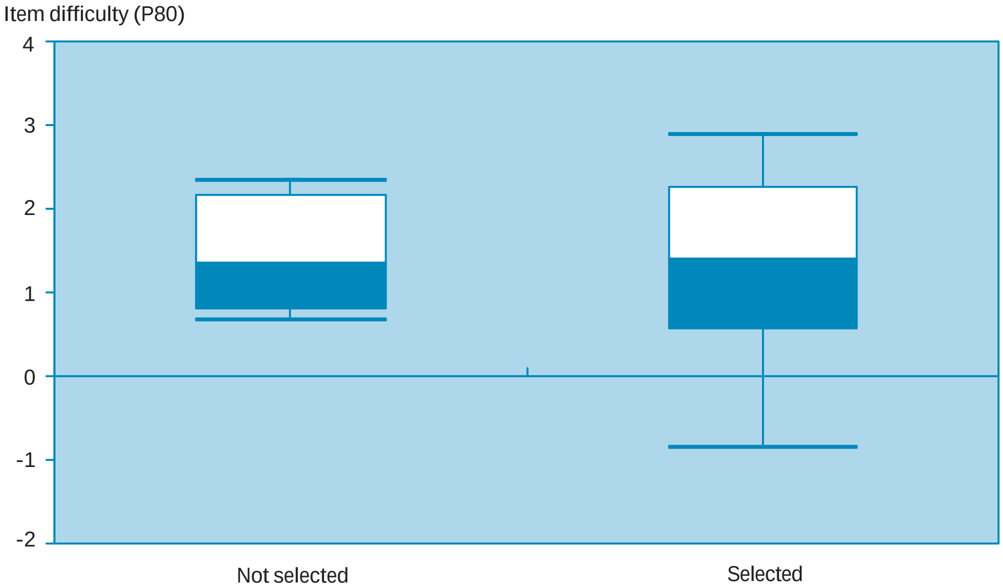
The final item selection was based on a number of criteria. First of all, the selected items were required to have good psychometric properties. Three parameters were focused on here: Difficulty, discrimination, and the item fit within the IRT-model. Items with extreme difficulty, weak discrimination properties and/or inadequate item fit within the IRT-model were eliminated. Secondly, it was crucial that the final set of selected items measured all proficiency levels of problem solving with the best possible distribution here as well. That also means that the overall range of difficulties was to be as large as possible and an even distribution of difficulties was aimed for. Third of all, the selected items were not supposed to show critical differences between national samples, unless of course these differences could be pinpointed to an operational error in the critical countries' material or procedure. Furthermore, some more technical aspects were also taken into account. It was attempted to keep the item types and question formats as balanced as possible. Time constraints were examined and adhered to, linguistic properties of the items were checked, and scoring problems resolved. It was also checked that the resulting projects still "made sense" from the point of view of the project's story or context.

Based on these criteria, 20 items were finally selected for the main study.

Figure 4 shows the distribution of item difficulties for those items that were selected (right) and those that were eliminated (left). As can be seen, the final version of the ALL problem-solving test covers the whole range of difficulties, with the exception of only one very difficult item. Mean difficulty was preserved from the pilot test versions to the final version of the test to be implemented in the main survey.

Figure 4

Item difficulty distributions for selected vs. eliminated items



Based on the results from Canada, Italy, Norway, and Switzerland, it can be shown that the selected items perform in a similar way in all countries. Some very few country by item interactions were identified, but in all cases the reasons could be traced back to layout, translation or adaptation mistakes. Also, the selected items make up short versions of projects with a realistic and suitable scenario, i.e. the items fit together well and the projects are plausible as a whole. Thus, it can be expected that the final version of the ALL problem-solving scale will show excellent psychometric properties in the main study and allow for a well-founded discrimination between the levels of analytical problem-solving proficiency. An attempt to identify those levels based on the selected items is presented in the next section.

5.5 Proficiency levels

5.5.1 Criteria and expectations

In chapter 2, four levels of analytical problem solving were defined. From a theoretical point of view, a scale covering the following competency levels was expected:

1. Content related reasoning
2. Evaluating
3. Ordering/Integrating
4. Critical Thinking

Experts classified each of the 20 selected items according to these categories. It was hypothesized that items classified as covering higher levels of proficiency should exhibit higher indices of item difficulty in the pilot test. However, as the empirical difficulty of a test item is shaped by a multitude of factors which can only partially be controlled for (e.g. the amount of previous knowledge required, the clarity of the item text, the mental workload involved, etc.), a certain amount of overlap between the predefined sets of items is inevitable. Previous work on proficiency scaling (cf. Watermann and Klieme, 2002) shows that sophisticated theories of item difficulty, operationalized by expert ratings of item demands, can explain between 65 and 80 percent of between-item-variance in item difficulty, when applied to large-scale assessment data.

The following two criteria are therefore realistic and should yield a satisfactory level of precision:

(a) Mean item difficulty should increase from level (1) to level (4).

(b) At least two thirds of the between-item-variance in difficulty should be explained by the experts' classification of items into the four proficiency levels.

5.5.2 Operationalizing the proficiency levels: The item classification

Level 1: 3 out of 20 items were classified as content-related tasks. These are rather concrete tasks with a limited scope of reasoning. They require the respondent to make simple connections, without having to check the constraints systematically. The respondent has to draw direct consequences, based on the information given and on her previous, content-related knowledge.

Thus, the mental operations that must be applied successfully to solve items at level 1 can be characterized as *schemata of content-related thinking*.

Level 2: Another 3 items were classified as corresponding to the second level. These items require the respondent to evaluate certain alternatives with regard to well-defined, transparent, explicitly stated criteria. The reasoning may be done step by step, in a linear process, combining information from the question section and the information section.

Thus, the mental operations that must be applied successfully to solve items at level 2 can be characterized as *systematical (concrete logical) reasoning*.

Level 3: 8 out of 20 items were classified as belonging this level. Some tasks require the respondent to order several objects according to given criteria. Others require her to determine a sequence of actions/events or to construct a solution by taking non-transparent or multiple interdependent constraints into account. This means that on level 3 the respondent has to cope with multi-dimensional or ill-defined goals.

Thus, the mental operations that must be applied successfully to solve items at level 3 can be characterized as *formal operations*. The reasoning process goes back and forth in a non-linear manner, requiring a good deal of self-regulation.

Level 4: The remaining 6 of the 20 items correspond to this level. These items require the respondent to judge the completeness, consistency and/or dependency among multiple criteria. In many cases, she has to explain how the solution was reached and why it is correct. The respondent has to reason from a "meta-perspective", grasping an entire system of problem states and possible solutions.

Thus, the mental operations that must be applied successfully to solve items at level 3 can be characterized as *critical thinking and meta-cognition*.

5.5.3 Findings and conclusions

Figure 5

Item difficulty parameters for the four classes of items

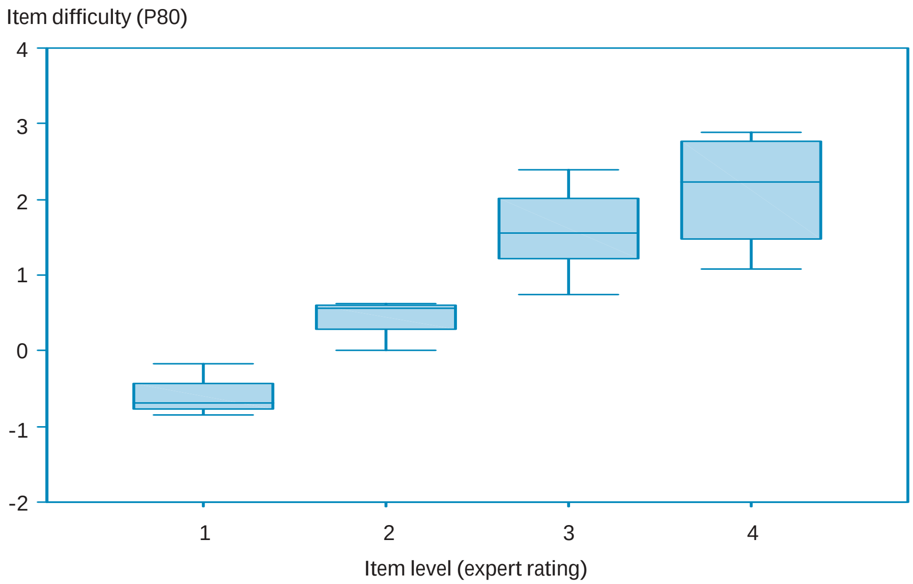


Figure 5 shows the distribution of item difficulty parameters within the four classes of items. For each level, the range from lowest to highest item difficulty is shown. Half of the item difficulty parameters are within the shaded area. The straight line at the center of the box indicates the median of item difficulties for the respective level. As expected, item difficulty shows a continuous increase from level 1 to level 4. Also, despite some overlap, the four ranges of item difficulty can be clearly discriminated. In fact, 77 percent of between-item variance can be attributed to the classification scheme. From these findings we can conclude that the ALL problem-solving scale can in fact be interpreted as a scale that measures the quality of analytical reasoning and decision making. According to the complexity of the required reasoning processes, the items can be classified into four levels as described above. The pilot results support this four-level model of item complexity, because the theoretically defined levels show a clear pattern of increasing empirical item difficulty.

When it comes to assessing the problem-solving competency of the respondents, it is possible to distinguish four levels of problem-solving proficiency analogous to the four levels of item complexity/difficulty. Respondents on level 1 can be expected to cope successfully only with content-related, concrete problems. At level 2, respondents are also able to evaluate actions and make decisions with regard to well-defined criteria. At level 3, respondents can apply formal operations which are needed to order and integrate multiple or ill-defined criteria. At the highest level, level 4, respondents are also able to evaluate a system of goals or criteria as a whole, applying critical thinking and meta-cognition.

5.6 Relation to other domains

To what extent does analytical problem solving in real world contexts, when measured with the project approach, overlap or go beyond literacy and numeracy skills? Reading skills, particularly document literacy, and some rudimentary numeracy skills, are necessary pre-requisites for successfully responding to the problem-solving tasks. We therefore expect the problem-solving score to co-vary with the scores of literacy and to a certain extent the numeracy score. Furthermore, we expect that analytical intelligence, in the psychometric sense (which is not measured in the ALL study), will influence the score.

The project tasks present several special aspects that go beyond literacy tasks and abstract intelligence tasks:

a) They are embedded in a complex, realistic, multi-step action context. Whereas the understanding of textual and document information is central for the literacy items, in problem-solving tasks the focus is on using this information in order to solve everyday-type tasks.

b) They demonstrate a high complexity made possible, for example, by integrating various representational formats (verbal, numerical, pictorial, diagrammatic), by setting multi-dimensional or ill-defined goals and by some intransparency in defining the problem situation. They therefore require more regulation activity than the more "transparent" literacy and numeracy tasks.

c) They require various kinds of reasoning activity, including analogical, inductive, deductive, and critical reasoning, involving content-related as well as logical and formal operations. As explained in section 2.5, the kind of reasoning required determines the complexity and difficulty of a problem.

These distinctive features of cross-curricular problem solving competency have also been highlighted in other assessment frameworks such as PISA 2003 (OECD, in press), PISA/Germany (Klieme et al. 2001) or the large-scale study in Hamburg/ Germany (Ebach, Klieme and Hensgen 2000). The empirical results of the two German projects clearly support the hypothesis that problem-solving tasks form a new dimension within the structure of competencies that is usually addressed by large-scale assessments. Based on structural equation modeling, it has been shown that problem-solving tasks of the "project" type do constitute a unique factor, which can be discriminated from reading literacy, numeracy/mathematical literacy, and also from psychometric intelligence.

When the ALL problem-solving items were developed, there was a clear intent to keep the language structure simple and straightforward. However, the frequently higher complexity of project tasks clearly requires a sufficient understanding of the instructions by the respondents, which itself pre-supposes a minimal level of reading literacy. This can result in respondents with very low literacy skills not understanding the project tasks, and, therefore, being unable to solve them. Thus, literacy skills are likely to have a threshold function for the analytical problem-solving scale. On the other hand, each of the literacy and numeracy tasks presented in ALL requires a certain amount of reasoning. Therefore, it could also be expected that a minimal level of analytical problem-solving proficiency is a prerequisite for prose and document literacy and numeracy. These competing hypotheses will be tested within the ALL main survey.

6. Conclusions

Based on research in the field of problem solving, this framework presents an assessment rationale for a meaningful subset of problem-solving skills, i.e. analytical problem-solving skills. Based on this rationale a set of assessment instruments was developed and further refined in the development process outlined in chapter 4, ranging from first draft scenarios to well-established projects which proved to “work” in a multinational pilot study. The assessment yields one scale for problem-solving skills with four competency levels:

1. Content related reasoning
2. Evaluating
3. Ordering/Integrating
4. Critical Thinking

The empirical results of the pilot study show conclusive results: All tested projects yielded very good empirical results and, with slight modifications, may be used in the main survey. The items within the projects cover a broad range of difficulties as well as all the competency levels. The long and short versions of the problem-solving projects proved to be diagnostically equivalent. Due to financial and logistical reasons, the short versions of the problem-solving projects – somewhat extended and modified – were selected for use in the ALL main survey.

References

Anderson, J.R. (1999). *Cognitive psychology and its implications*. 5th edition. Cambridge, Mass.: MIT Press.

Baker, E.L. (1998). *Final report for validation of teamwork skills questionnaire using computer-based teamwork simulations* (85300-97-0056). Ottawa, Canada: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baxter, G.P., and Glaser, R. (1997). *An approach to analyzing the cognitive complexitiy of science performance assessments* (CSE Technical Report 452). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Binkley, M.R., Sternberg, R., Jones, S., and Nohara, D. (1999). *An overarching framework for understanding and assessing life skills*. Unpublished International Life Skills Survey (ILSS) Frameworks, Washington.

Bloom, B.S., Hastings, J.T., and Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.

Boshuizen, H.P.A., van der Vleuten, C.P.M., Schmidt, H.G., and Machiels-Bongaerts, M. (1997). Measuring knowledge and clinical reasoning skills in a problem-based curriculum. *Medical Education*, *31*, 115-121.

Brown, A.L. (1987). Metacognition executive control, self-regulation and other more mysterious mecanism. In F.E. Weinert and R.H. Kluwe (Eds.), *Metacognition, Motivation and Understanding*. Hillsdale, NJ.: LEA.

Buchner, A., and Funke, J. (1993). Finite State Automata: Dynamic Task Environments in Problem Solving Research. *Quaterly Journal of Experimental Psychology, 46A*, 83-118.

Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York: Cambridge University Press.

Charles, R., Lester, F., and O'Daffer, P. (1987). *How to evaluate progress in problem solving*. (Vol. 6): National Council of Teachers of Mathematics.

Chi, M.T.H., Glaser, R., and Farr, M.J. (Eds.). (1988). *The nature of expertise*. Hilldale, NJ.: Erlbaum.

Collis, K.F., Romberg, T.A., and Jurdak, M.E. (1986). A technique for assessing mathematical problem-solving ability. *Journal for Research in Mathematics Education*, *17*(3), 206-221.

Didi, H.-J., Fay, E., Kloft, C., and Vogt, H. (1993). *Einschätzung von Schlüsselqualifikationen aus psychologischer Perspektive* (Gutachten im Auftrag des Bundesinstituts für Berufsbildung (BIBB)). Bonn: IBF - Institut für Bildungsforschung.

Dörner, D., Kreuzig, H., Reither, F., and Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58, Whole No. 270.

Ebach, J., Hensgen, A., Klieme, E., and Sidiropoulou, E. (1999). Analytical reasoning in real world contexts (pp. 1-15). Luxembourg: International Life Skills Survey (ILSS).

Ebach, J., Klieme, E. and Hensgen, A. (2000). *Der SL HAM 6/7 Problemlösetest*. (Report on a large scale study of problem solving in grades 6/7 in Hamburg/ Germany.) Unpublished report. Bonn: Institut für Bildungsforschung.

Ennis, R.H. (1996). *Critical Thinking*. Upper Saddle River, NJ 07458: Prentice Hall.

Flavell, J.H. (1976). Metacognitive aspects of Problem solving. In T.n.o. intelligence (Ed.), *Hillsdale*, NJ. Erlbaum.

Frensch, P.A., and Funke, J. (Eds.). (1995). *Complex problem solving. The european perspective*. Hillsdale, NJ.: Lawrence Erlbaum Ass.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. Sternberg and P.A. Frensch (Eds.), *Complex Problem Solving: Principles and Mechanisms* (pp. 185-222). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Funke, J., and Fritz, A. (Eds.). (1995). *Neue Konzepte und Instrumente zur Planungsdiagnostik*. Bonn: Deutscher Psychologen Verlag.

Hiebert, J., Carpenter, T.P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., and Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: the case of mathematics. *Educational Researcher, 25*(4), 12-21.

Hunt, E. (1994). Problem solving. In R.J. Sternberg (Ed.), *Thinking and Problem Solving* (pp. 215-232). San Diego: Academic Press.

Jeck, S. (1997). *Planen und Lösen von Alltagsproblemen*. Lengerich: Pabst.

Klieme, E. (1989). Mathematisches Problemlösen als Testleistung. Frankfurt/M: Lang.

Klieme, E., Ebach, J., Didi, H.-J., Hensgen, A., Heilmann, K., and Meisters, K.-H. (in press). Problemlösetest für die 7. Jahrgangsstufe. In R. Lehmann and I. Steinbach (Eds.), Hamburger Schulleistungstest für sechste und siebente Klassen.Göttingen: Hogrefe.

Klieme, E., Funke, J., Leutner, D., Reimann, P., and Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Ergebnisse aus einer Schulleistungsstudie. Zeitschrift für Pädagogik, No. 1/2001

Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. Ingenkamp, and R.S. Jäger (Eds.), *Tests und Trends - Jahrbuch der Pädagogischen Diagnostik, Bd. 10* (pp. 11-39). Weinheim: Beltz.

Lave, J. (1988). *Cognition in practice*. Boston, MA.

Leutner, D. (1992). *Adaptive Lehrsysteme*. Weinheim: Beltz/PVU.

Leutner, D. (1999). *Discovery learning, intelligence, and problem solving ability.* Unpublished Paper presented at the EARLI Conference. Gothenburg, September 1999, Erfurt Educational University, Erfurt.

Luchins, A. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54* (6, Whole No. 248).

Mayer, R.E. (1992). *Thinking, problem solving, cognition*. (2 ed.). New York: W.H. Freeman and Company.

Mayer, R.E., and Wittrock, M.C. (1996). Problem-solving transfer. In D.C. Berliner and R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Simon and Schuster Macmillan.

Meijer, J., and Elshout-Mohr, M. (1999). *An instrument for the assessment of cross curricular skills*. Unpublished Paper presented at the EARLI conference, Göteborg, Schweden, 1999.

Mumford, M.D., Supinski, E.P., Baughman, W.A., Costanza, D.P., and Threlfall, K.V. (1997). Process-based measures of creative problem-solving skills: V. overall prediction. *Creativity Research Journal, 10*(1), 73-85.

Neisser, U., Boodoo, G., Bouchard Jr., T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., and Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist, 51*, 77-101.

Newell, A., and Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs: Prentice-Hall.

Norris, S.P. (1989). Evaluating critical thinking ability. *History and social science teacher, 21*(3), 135-146.

O'Neil, H. (1999). *A theoretical basis for assessment of problem solving.* Unpublished Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 19, 1999, University of Southern California.

O'Neil, H.F.J., Chung, K.W.K.G., and Herl, H.E. (1999). *Computer-based collaborative knowledge mapping to measure team processes and team outcomes.* University of Southern California, Los Angeles.

OECD. (1997). *Prepared for life? How to measure cross-curricular competencies*. Paris: OECD.

OECD. (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD.

OECD (in Press). *PISA 2003 framework for the assessment of problem solving skills*. Paris: OECD.

Pólya, G. (1945, 1980). *How to solve it. Deutsche Übersetzung: Schule des Denkens. Vom Lösen mathematischer Probleme*. (Vol. 3). Bern: Francke.

Raaheim, K. (1988). Intelligence and task novelty. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 73-97).

Reimann, P., and Schult, T.J. (1996). Turning Examples Into Cases: Acquiring Knowledge Structures for Analogical Problem Solving. *Educational Psychologist, 3*(2), 123-132.

Ryjchen, D. and Salganik, L.H. (2000). *Definition and selection of key competencies (DeSeCo)*. Paris, France: OECD.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, Metacognition, and Sense Making in Mathematics. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 334-370).

Segers, M.S.R. (1997). An alternative for assessing problem-solving skills: the overall test. *Studies in Educational Evaluation*, 23(4), 373-398.

Smith, M. (1991). *Toward a Unified Theory of Problem Solving*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sternberg, R.J. and Wager, R.K. (1986). *Practical Intelligence*. Cambridge: Cambridge University Press.

Sternberg, R.J., and Frensch, P.A. (Eds.). (1991). *Complex problem solving. Principles and mechanism*. Hillsdale, NJ.: Lawrence Erlbaum Ass.

Sternberg, R.J., and Kaufman, J.C. (1998). Human Abilities. *Annual Review of Psychology, 49*, 479-502.

Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau, 50*(4), 220-228.

Svecnik, E. (1999). *Stellenwert der Problemlösefähigkeit im österreichischen Schulwesen.* Unpublished Bericht des EU-Netzwerk "New Assessment Tools for Cross-Curricular Competencies in the Domain of Problem Solving", Graz.

Swaak, J., and de Jong, T. (1996). Measuring intuitive knowledge in science: the development of the what-if test. *Studies in Educational Evaluation, 22*(4), 341-362.

Trier, U.P., and Peschar, J. (1995). Cross-Curricular Competencies: Rationale and Strategy for Developing a New Indicator. In OECD (Ed.), *Measuring what Students Learn* (pp. 99-109). Paris: OECD.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Watermann, R. and Klieme, E. (2002). Reporting results of large-scale assessment in psychologically and educationally meaningful terms - Construct validation and proficiency scaling in TIMSS. *European Journal of Psychological Assessment*, *18*, 190-203.

Weinert, F.E. (1999). *Konzepte der Kompetenz.* Unveröffentlichtes Gutachten zum OECD Projekt "Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo). München.

Weinert, F.E., and Kluwe, R.H. (Eds.). (1987). *Metacognition, Motivation and Understanding*. Hillsdale, NJ.: LEA.

Zsambok, C.E., and Klein, G. (Eds.). (1997) *Naturalistic Decision Making*: Mahwah, N.J.: Lawrence Erlbaum Ass.

Part III

Theoretical Frameworks for Specific Domains Not Included in ALL and for Which Assessment Tools Remains to be Developped

This part of the publication includes three chapters. Chapters 7 and 8 provide assessment frameworks for two skill domains where development failed to yield approaches to measurement that were sufficiently robust to meet the criteria set out for inclusion in the international comparative assessment. These chapters also set out what was learned during the process of development and validation. Chapter 7 presents the framework for teamwork and chapter 8 presents the framework for practical cognition. Chapter 9 provides the assessment framework developed by the Educational Testing Service (ETS) for measuring information and communication technology literacy. Although it was not available soon enough to inform the design of the ALL study, pilot testing has revealed that the framework is viable and the approach to measurement yields robust estimates that would meet the demanding criteria set for ALL.

Chapter 7

The ALL Teamwork Framework

David P. Baker	*American Institutes for Research, U.S.*
Lisa Horvath	*George Washington University, U.S.*
Michael Campion,	*Purdue University, U.S.*
Lynn Offermann	*George Washington University, U.S.*
Eduardo Salas	*Orlando, FL, U.S.*

Table of Contents

Table of Contents

Summary

Governments, businesses, and community groups are increasingly relying on work teams to streamline processes, enhance participation, and improve performance. Teamwork is of worldwide importance; individuals who wish to participate fully in community and professional life must increasingly possess the skills necessary to work in teams.

Although teams are diverse and can take on many forms, all teams are defined by four characteristics. They have two or more individuals; they share a common goal(s); they are task-interdependent; and they have a desired productive outcome(s). These characteristics serve as the basis for developing a working definition of a "team," a definition that the ALL can use to provide insight regarding the prevalence and the expression of teamwork skills across various cultures.

The Teamwork scale of ALL seeks to assess the core skills associated with teamwork. To this end, three primary skills required for effective teamwork—Group Decision Making/Planning, Adaptability/Flexibility, and Interpersonal Relations—are proposed, each represented by distinct behavioral manifestations. Group Decision Making/Planning refers to the ability to identify problems and gather, evaluate, share and link information. Adaptability/Flexibility implies using a variety of task-relevant strategies, providing assistance, adjusting to task reallocation and accepting feedback. Interpersonal Relations reflects supporting team decisions, sharing work, helping others, and seeking mutually agreeable solutions. Communication skills—including providing complete and concise information, listening effectively, and asking questions—underlie the other three skills and serve as a bridge among them. In addition, two other factors play key roles in teamwork: attitudes toward teamwork and past experience with teams.

By definition, teamwork skills can only be observed directly in a teamwork setting. However, because direct observation is not consistent with ALL methodology, respondent teamwork skills should be assessed indirectly. Specifically, knowledge of teamwork skills, attitudes towards teamwork, and past experience in teams should be measured, and links should be drawn between these constructs and team performance.

Finally, teamwork, more than other life skills, is likely to be affected by culture. Although the team skills described in this framework are assumed to define teamwork generally, the behavioral manifestation of these skills is likely to vary across cultures. Respondent performance should be interpreted relative to the effective teamwork behaviors defined for a given country, thereby providing information regarding national attitudes toward teamwork and regarding the degree to which behavioral expressions of teamwork skills vary across nations. This information can be used by employers and educators alike, to assess and improve teamwork in a nation's workforce and general population.

1. Introduction

This document presents a framework for assessing Teamwork as part of the Adult Literacy and Life skills survey (ALL). The framework was developed from the literature on teams and what is currently known about teamwork. Overall, the framework serves three purposes. First, it bounds the problem domain by clearly specifying the critical components of teamwork to assess. Our goal here is to target the most fundamental aspects of teamwork. Second, the framework drives our approach to measurement. Strategies that are most effective for assessing team knowledge, skills, and attitudes should be identified and selected. Finally, based on the measurement strategies identified, the framework serves as the template for item development. Items should be developed to target key aspects of teamwork that are specified in this framework.

The framework is divided into five sections. The first presents a detailed discussion of the literature on teams and what is currently known about the knowledge, skills, and attitudes required for effective team performance. Here, we present our definition of a team and clearly delineate the core facets of teamwork. Rather than including all variables, we present the core dimensions that characterize what teams do. These dimensions are assumed to be central to all teams, regardless of culture.

In the second section, we draw on the results of our literature review to build a model of teamwork. The purpose of this model is to identify key areas for measurement in ALL. In particular, we propose that the teamwork measure should assess what team members bring to a team (e.g., attitudes, past experience, etc.) and what team members do in a team (e.g., interact, coordinate, etc.).

Our proposed framework is intended to target the most fundamental aspects of teamwork and recognizes that the primary goal of the teamwork measure is not to assess differences in culture. Nonetheless, given the interpersonal nature of teamwork, we anticipate cultural differences. The third section of this framework addresses this issue. In particular, relevant cultural research is reviewed, and the implications of these studies are discussed in light of our objective (i.e., measuring team knowledge, skills, and attitudes internationally).

Once the key facets of teamwork are identified and the possible effects of culture are discussed, the fourth section of this framework presents specific strategies for measuring teamwork. We first present the theoretical and practical assumptions that guide our approach. Next, we describe each proposed teamwork measure with respect to the measurement approach employed, the process by which items were developed, and the procedures for scoring.

Finally, the fifth section of this framework briefly discusses social and economic indicators that may affect teamwork. Here, variables are proposed for the respondent background questionnaire. Information on these variables, which are expected to moderate participant responses on the teamwork measure, should also provide insights into the determinants of teamwork in different nations.

1.1 Why measure teamwork?

Organizations (both work and non-work) are increasingly using teams to streamline processes, enhance participation, and improve quality (Cohen and Bailey, 1997). Hence, teams are becoming the primary building block of most organizations (Brooks, 1993; McGrath, 1997). In fact, a recent study by Gordon (1992) found that 82% of U.S. companies with 100 or more employees utilize some form of teams. Teams are found in such diverse fields as education, religion, science, manufacturing, and consulting.

Because teams span both private and public life, individuals must be able to work and perform in a team context to function effectively in today's society. Both the Secretary's Commission on Achieving Necessary Skills (SCANS; U.S. Department of Labor, 1991, 1992a, 1992b) and the Conference Board of Canada Employability Skills Profile (1993) cite the importance of interpersonal skills (or teamwork) in work and everyday life.

Due to its prevalence in society, teamwork has been identified as an important life skill. Consistent with the goals of ALL, the teamwork measure should provide information as to how teamwork skills are distributed in the adult population internationally. Information on the nature of teamwork skills associated with a particular nation and the social and economic factors that influence the development of teamwork skills are of particular interest. This information should prove valuable to employers and educators who wish to improve teamwork in the workforce and elsewhere.

1.2 Challenges of the project

Although there is little doubt that teamwork is an important life skill, the measurement of teamwork in ALL presents specific challenges. First, ALL will be the initial attempt to provide a large-scale international assessment of teamwork skills. As a result, a limited number of methods and approaches exist as precedents. Past international assessments have focused on adult literacy (i.e., see IALS) as opposed to interpersonal skills like teamwork. Therefore, we expect to learn a great deal about teamwork skills and their distribution in the adult population across nations.

Second, unlike other life skills measured by ALL, teamwork will likely be affected by culture. Although we believe that a certain set of core skills defines teamwork across all cultures, the way in which these skills are manifested within a team is likely to vary. Therefore, we will not attempt to develop an invariant set of items to be translated for use in each nation. Rather, items should be modified as necessary to take known cultural differences into account.

Finally, and perhaps most challenging, is the fact that most methods of assessing teamwork skills require direct observation of team performance (D. Baker and Salas, 1992; 1997; Brannick, Prince and Salas, 1997; Ilgen, 1999). Typically, team members are placed in a scenario. Experts observe team behaviors and provide performance ratings on specific teamwork skills. This measurement approach differs substantially from the approach to be used in ALL. Here, the teamwork measure will be a short paper-based instrument; no opportunity for directly observing the teamwork skills of the respondent will be available. Thus, respondents' teamwork skills will have to be assessed indirectly instead of directly. For ALL, we propose to measure the knowledge of teamwork skills, attitude towards teamwork, and past experience in teams and then attempt to draw links between these variables and team performance.

With these challenges in mind, we turn to a discussion of the key components of teamwork underlying our framework. Because these domains will drive development of the teamwork measures for ALL, we draw heavily from the literature on teams and on what is currently known about teamwork.

2. Teams and teamwork

2.1 What is a team?

Although a widespread consensus acknowledges the prevalence of teams in society, the research literature reflects only marginal agreement concerning the definitional components of teams. The variance in definitions is due in part to the diversity of team types. Teams carry a variety of purposes (e.g., learning, producing a product, solving problems, gaining acceptance), forms (e.g., virtual, co-located), and sizes and longevity (e.g., adhoc, long term) (Cohen and Bailey, 1997).

In an attempt to extract key features of teams and develop a working definition of teams for ALL, we reviewed several often-cited definitions (Dyer, 1984; Guzzo and Shea, 1992; Mohrman, Cohen, and Mohrman, 1995; Salas, Dickinson, Converse and Tannenbaum, 1992). This process produced four common characteristics of a "team."

- Two or more individuals
- A shared or common goal(s)
- Task interdependency
- A desired productive outcome(s)

These characteristics serve as the basis for developing our working definition of a "team." A clear definition of a team is essential because it provides measurement boundaries and clearly distinguishes teams from small groups, which do not necessarily connote interdependence. (A team is also a "small group," but a small group may or may not be a team.) Our definition of a team is as follows:

> *A team consists of two or more individuals who must interact to achieve one or more common goals that are directed toward the accomplishment of a productive outcome(s).*

In addition, the definition and core characteristics provide preliminary insight into the nature of teamwork and its key facets. For example, the characteristics of task interdependency and shared goals imply that team members must collectively decide on team goals (team decision making) and work cooperatively (coordination) to achieve these goals.

2.2 What is teamwork?

Teamwork has traditionally been described in terms of classical systems theory in which team inputs, team processes, and team outputs are arrayed over time. Here, team inputs include the characteristics of the task to be performed, the elements of the context in which teamwork occurs, and the attitudes team members bring to a team situation. Team process includes the interaction and coordination among members required for performing team tasks and achieving specific goals. Team outputs consist of the products that result from team performance (Hackman, 1987; Ilgen, 1999; McGrath, 1984). With regard to teamwork, the process phase is the defining point at which teamwork occurs; it is during this phase that team members interact and work together to produce team outputs.

Numerous theories have been proposed and extensive research has been conducted on the nature of team process (i.e., teamwork). Historically, this literature has sought to identify generic teamwork skills that are associated with most teams. More recently, the focus has shifted towards researchers identifying the specific *competency requirements* of team members (Cannon-Bowers, Tannenbaum, Salas, and Volpe, 1995; O'Neil, Chung, and Brown, 1997; Stevens and Campion, 1994). The term *competency* has a variety of meanings. However, it is generally used to denote the qualities needed by a jobholder (Boyatzis, 1982)[1]. Specifically, Parry (1998) defined the term "competencies" as a cluster of related knowledge, skills, and attitudes that affects a major part of one's job (i.e., one or more key roles or responsibilities); is correlated with performance on the job; can be measured against well-accepted standards; and can be improved through training and development.

Regarding teamwork, team competencies are the qualities needed by team members. Cannon-Bowers et al. (1995) identified three types of competencies that are central for effective teamwork: (1) team knowledge competencies, (2) team skill competencies, and (3) team attitude competencies.

Team Knowledge Competencies. Team knowledge competencies are defined by Cannon-Bowers et al. (1995) as the principles and concepts that underlie a team's effective task performance. To function effectively in a team, team members must know what team skills are required, when particular team behaviors are appropriate, and how these skills should be utilized in a team setting. In addition, team members should know the team's mission and goals and be aware of each other's roles and responsibilities in achieving those goals. Such knowledge enables team members to form appropriate strategies for interaction, to coordinate with other team members, and to achieve maximum team performance.

Team Skill Competencies. Team skill competencies, which have received considerable research attention, are defined as a learned capacity to interact with other team members at some minimal proficiency level (Cannon-Bowers et al., 1995). However, Cannon-Bowers et al. has reported that the literature on team skills is confusing and contradictory, as well as plagued with inconsistencies in terms of both skill labels and definitions. Across studies, different labels are used to refer to the same teamwork skills or the same labels are used to refer to different skills. In an attempt to resolve these inconsistencies, Cannon-Bowers et al., found that 130 skill labels could be sorted into eight major teamwork skill categories: adaptability, situation awareness, performance monitoring/ feedback, leadership, interpersonal relations, coordination, communication, and decision making. Numerous investigations have shown that these skills are directly related to team performance (see for example, Morgan, Glickman, Woodward, Blaiwes, and Salas, 1986; Oser, McCallum, Salas, and Morgan, 1992; Salas, Bowers, and Cannon-Bowers, 1995; Salas, Fowlkes, Stout, Milanovich, and Prince, 1999).

Team Attitude Competencies. Team attitude competencies are defined as an internal state that influences a team member's choices or decisions to act in a particular way (Cannon-Bowers et al., 1995; Dick and Carey, 1990). Attitudes toward teamwork can have a significant effect on how teamwork skills are actually put into practice. Positive attitudes toward teamwork and mutual trust among team members are examples of critical attitudes related to team process (Gregorich, Helmreich and Wilhelm, 1990; Ruffell-Smith, 1979; Helmreich, Fushee, Benson, and Russini, 1986). For example, Vaziri, Lee, and Krieger (1988) found that higher levels of mutual trust among team members led to a more harmonious and productive team environment. Finally, an attraction to being part of a team (i.e., collective orientation) is critical (Eby and Dobbins, 1997). Driskell and Salas (1992) reported that collectively-oriented individuals performed significantly better than did individually-oriented team members because

1. Boyatzis (1982), in his seminal work on competencies, defines a job competency as "an underlying characteristic of a person, which results in effective or superior performance in a job."

collectively-oriented individuals tended to take advantage of the benefits offered by teamwork. Furthermore, collectively-oriented individuals had the capacity to take other team members' behavior into account and believed that a team approach was superior to an individual one.

Refining the work of Cannon-Bowers et al. (1995), Cannon-Bowers and Salas (1997) delineated three types of team knowledge, skills, and attitude competencies. First, "*individual* competencies" are defined as the knowledge, skills, and attitudes required on the part of individual team members to perform position requirements. These competencies enable team members to perform tasks that are specifically assigned to them. For example, an individual in a marketing team assigned to purchase newspaper-advertising needs to possess specific knowledge and skills to successfully perform this task. Second, "*team* competencies held at the *individual level*" are defined as the knowledge, skills, and attitudes that are generic with respect to a team and its tasks. Essentially, these competencies are transportable to different teams and different team settings. For example, knowledge about teamwork skills and behaviors; skill in communication, team decision making, and interpersonal relations; positive attitudes toward teamwork, and a collective orientation enable team members to function effectively across a wide variety of teams. Finally, "*team* competencies held at the *team level*" are defined as the knowledge, skills, and attitudes that are specific to a particular team and task. Unlike team competencies at the individual level, these competencies are not transportable. They only have meaning within the team. For example, knowledge of teammate roles and responsibilities and specific teammate characteristics are only useful within a specific team context.

Given that the primary goal of ALL is to assess teamwork in the adult international population, teamwork measures will assess "*team* competencies held at the *individual level*." By definition, these competencies are of great interest to policymakers and educators because they enable individuals to function effectively in a wide variety of teams and a wide variety of team settings.

2.3 Core team skills, knowledge, and attitudes

A comprehensive review of teamwork models and research was conducted (e.g., Carnevale, Gainer and Meltzer, 1990; Commission on the Skills of the American Workforce, 1990) to identify core team knowledge, skills, and attitude competencies held at the individual level. From that broad review, we selected the most comprehensive and current team competency models (Cannon-Bowers et al., 1995; O'Neil et al., 1997; Stevens and Campion, 1994a) and used these models to identify core team competencies to measure in ALL. Competencies were selected based upon the following criteria: (1) the competencies were held at the individual level; (2) at least two of the three models delineated the competency (in some form); and (3) empirical research supported a positive relationship between the competency and performance.

Core Team Skills. Team skill competencies are discussed first because they represent the manifest, individual-level behaviors that the ALL measure is designed to assess. Four competencies were identified as "core" team skills competencies: communication, interpersonal relations (which includes cooperation and dealing with conflict), group decision making/planning, and adaptability/flexibility. Team leadership, an often-cited skill competency (see for example, Cannon-Bowers et al., 1995), was not included because our current focus is on the ability to *work* in a team, not to *lead* one. Each core team skill is defined below, along with behavioral examples that typify the skill's expression. Although this core is assumed to reflect teamwork in most cultures, it should be noted that the behavioral exemplars presented here were derived from research conducted on teams in the U.S. (Cannon-Bowers et al., 1995; O'Neil et al., 1997; Stevens and Campion, 1994a). As such, they may or may not be consistent with the *expression* of the same core skills in other cultures. Thus, the cross-cultural generalizability of behaviors that manifest core team skills in the U.S. remains an empirical question that the ALL will address. However, should cultures to which these behaviors do not generalize be included in the ALL, other behaviors are expected to express the same core team skill competencies systematically.

Communication is defined as establishing effective communication between self and others; it involves the exchange of clear and accurate information and the ability to clarify or acknowledge the receipt of information.

Strong communication skills are demonstrated by team members who

- Provide clear and accurate information
- Listen effectively
- Ask questions
- Acknowledge requests for information
- Openly share ideas
- Attend to non-verbal behaviors

Interpersonal Relations is a broad area that encompasses cooperation and dealing with conflict within the team. Therefore, effective interpersonal relations include working cooperatively with others, working together as opposed to working separately or competitively, and resolving disputes among team members.

Strong interpersonal relations skills are demonstrated by team members who

- Share the work
- Seek mutually agreeable solutions
- Consider different ways of doing things
- Manage/Influence disputes

Group Decision Making/Planning is defined as the ability of a team to gather and integrate information, use logical and sound judgment, identify possible alternatives, select the best solution, and evaluate the consequences.

Strong group decision making and planning skills are demonstrated by team members who work with others to

- Identify problems
- Gather information
- Evaluate information
- Share information
- Understand decisions
- Set goals

Adaptability/Flexibility is defined as the process by which a team is able to use information gathered from the task environment to adjust strategies through the use of compensatory behavior and reallocation of intra-team resources.

Strong adaptability/flexibility skills are demonstrated by team members who

- Provide assistance
- Reallocate tasks
- Provide/Accept feedback
- Monitor/Adjust performance

Core Knowledge Competencies. Regarding the core knowledge competencies, team members must know how and when to use the teamwork skills listed above. Therefore, team knowledge competencies include knowing how to communicate with other team members, how to interact and resolve conflicts, how to plan and make team decisions, and how to adapt and provide assistance to other team members. Such knowledge enables individuals to execute critical teamwork skills and function effectively in a team environment.

The core team knowledge competencies identified above are considered as prerequisites to skill execution. These knowledge competencies are critical components of each team skill (i.e., they comprise the knowledge part of the skill). We present them separately to distinguish what we believe are two critical facets of teamwork: knowing what to do in a team versus doing it. Although the ALL measure focuses on the behavioral alternatives respondents choose in team situations, we believe that knowledge competencies, as defined, are directly related to team member skills and to the level of teamwork achieved.

Core Attitude Competencies. Finally, two attitude competencies were identified: Belief in the Importance of Teamwork and Collective Orientation. These attitudes are brought to the team setting by individuals and can influence the nature of teamwork within a team. As Driskell and Salas (1992) point out, individuals who tend to possess positive attitudes toward teamwork are most likely to take advantage of the benefits teamwork has to offer. Such individuals believe a team approach is better than an individual one; compared to individually-oriented team members, they are better at taking another team member's behavior into account. Each attitude competency is briefly defined below.

Belief in the Importance of Teamwork is defined as the belief that teamwork is critical for successful performance of team tasks.

Collective Orientation is defined as an attraction to, or desire to be part of, a team.

3. A model of teamwork

Based on the literature review and what is generally known about teamwork, Figure 1 presents a model for understanding teamwork for the purposes of ALL. Referring to Figure 1, several things should be noted. First, the skill competencies of Group Decision Making/Planning, Adaptability/Flexibility, and Interpersonal Relations are at the core of teamwork. We believe that team members must know how and when to use these competencies to function effectively within the team. Second, we propose that Communication spans each of the three core areas; it is the glue that holds the team together. For example, Group Decision Making/Planning cannot be accomplished within a team unless team members provide clear and accurate information, listen effectively, and ask questions. Finally, the model proposes that the extent to which an individual is drawn toward teamwork, believes in the importance of teamwork, and has experienced team activity will influence how effectively team skills and behaviors are executed.

Figure 1 also presents a starting point for developing measures for ALL by identifying specific variables to be measured. These include the skills of Group Decision Making/Planning, Adaptability/Flexibility, Interpersonal Relations and Communication, and the attitudes Belief in the Importance of Teamwork and Collective Orientation. Furthermore, Figure 1 presents specific behavioral examples of each skill, as discussed above. These behavioral indicators will be used to construct responses for items measuring teamwork skills. Items that tap respondents' belief in the importance of teamwork and their collective orientation will also be included in the ALL measure.

Figure 1

ALL model for understanding teamwork

Attitudes and experience	Skills		
Attitudes and dispositions	**Group Decision Making/Planning**	**Adaptability/ Flexibility**	**Interpersonal Relations**
	Identify problems	Provide assistance	Share the work
	Gather information	Reallocate tasks	Seek mutually agreeable solutions
Experiences	Evaluate information	Provide/Accept feedback	Consider different ways of doing things
	Share information	Monitor/Adjust performance	Manage/Influence disputes
Implicit theories about teamwork	Understand decisions		
	Set goals		
	Communication Provide clear and accurate information Listen effectively Ask questions Acknowledge requests for information Openly share ideasPay attention to non-verbal behaviors		

Prior to discussing our method and approach for developing the Teamwork Scale for ALL, we briefly review the relevant literature on culture. More than other ALL measures, responses to the teamwork measure may be affected by the culture of the respondent. In the next section, we review research that specifically examines the relationship between societal culture and an individual's attitudes, values, beliefs, and behavior in a team. Based on this research and on our understanding of the factors that enhance teamwork, we propose a number of likely relationships that will be demonstrated in the ALL between culture and teamwork.

4. Culture and teamwork

Culture is simply "*the values, beliefs, behavior, and material objects that constitute a people's way of life*" (Macionis, 1993). Research examining the relationship between culture and performance in organizations has tended to focus on people's attitudes, values, beliefs, sources of motivation, and satisfaction and is commonly assumed to predict behavior.

Although alternative categorizations exist (e.g., Trompenaars, 1993), the most commonly used description of cultural comparisons has been developed by Hofstede (1980; 1991). Hofstede conducted the most exhaustive cross-cultural study to date (questionnaire data from 80,000 IBM employees in 66 countries across seven occupations) and established four dimensions of national culture. The four dimensions are the following:

> Power Distance: The extent to which the less powerful members of institutions and organizations accept that power is distributed unequally.
>
> Individualism/Collectivism: The extent to which a society is a loosely knit social framework in which people are supposed to take care only of themselves and their immediate families, as opposed to tight social frameworks in which people are integrated into strong cohesive groups that look after them in exchange for loyalty.
>
> Uncertainty Avoidance: The extent to which people feel threatened by ambiguous situations and have created beliefs and institutions that try to avoid them.
>
> Masculinity/Femininity: The extent to which the dominant values in a society tend toward achievement and success and away from caring for others and quality of life.

Research has shown that social dynamics vary according to the norms individuals hold concerning appropriate social behavior and that these norms vary across cultural settings (Triandis, 1989). For example, direct confrontation of one's boss may be acceptable in one culture and avoided in another (Adler, 1986). In fact, preliminary empirical studies have demonstrated large cross-national differences in attitudes regarding task performance across several work domains (Hofstede, 1980; Merritt, 1996; Merritt and Helmreich, 1996). Therefore, it is reasonable to suspect that societal culture exerts important effects on team members' knowledge of acceptable team skills, on members' attitudes toward teamwork, and on team behavior.

4.1 Research on culture and teamwork

Several notable studies have examined the attitudinal differences among workers of different cultures (Evan, 1993). Hofstede (1985) explored a matched sample of employees in a single, multinational corporation in 40 countries. He found wide differences in attitudes toward collaboration. Individualistic countries were more likely to reject collaborative work, preferring to work on their own, whereas collectivist cultures preferred collaborating with others. In related work, Kelly and Reeser (1973) examined the differences between American managers of Japanese ancestry and those of Caucasian ancestry. Similarly, a study by Pizam and Reichel (1977) examined the differences between Israeli managers of Oriental ancestry and those of Western ancestry. In both studies, cultural differences were observed in areas such as respect for formal authority, commitment to long-term employment, paternalism with respect to subordinates, and interest in teamwork.

Cross-national differences in attitudes toward interpersonal interactions have also been found in aviation teams (Helmreich, Merritt, and Sherman, 1996). Current research has demonstrated substantial variability among cultures concerning attitudes toward command responsibility and the captain's role on the flight deck. Cultures differ with respect to members' belief that junior crew members should question the actions of captains. Similarly, individuals from different cultures differ significantly in their endorsement of whether or not they should speak up when they perceive a problem with the flight. Overall, Anglos are more likely than non-Anglos to believe that it is acceptable for crew members to question the captain's decisions, that it is acceptable for the first officer to assume command of the aircraft under certain circumstances, that the captain should not automatically take physical control, and that successful flight deck management depends more than on the captain's individual proficiency.

Parallel findings were found in cross-cultural research, conducted at the Center for Creative Leadership, on teamwork and team leadership. More judicious use of personal prominence and power, greater openness to the ideas and interest of others, and mitigation of tough mindedness are more acceptable among team leaders in Europe, as compared to those in the U.S. (Leslie and Van Velsor, 1998).

Finally, Gibson (1996) found that the relation between team beliefs and team performance differed between American and Indonesian work teams. A collective orientation enhanced team performance, whereas an individualistic orientation inhibited teamwork. Kirkman (1997) found that, in the U.S., Finland, Belgium, and the Philippines, the amount of resistance to working in a team varied, depending upon the cultural orientation of employees. Respondents with individualistic values resisted working in teams more than did respondents with collectivist values. Further, respondents who valued power distance reported higher levels of resistance to self-management than did those who placed a low value on power distance. Currently, Gibson and Zellmer (1997) are engaged in an intercultural analysis on the meaning of teamwork. Although their preliminary results demonstrate that teams have become a pervasive element across the world, the concept of teamwork itself seems to differ as a function of culture.

4.2 Implications for measuring teamwork internationally

Based on the research cited above, it appears that culture can significantly affect the way in which individuals communicate, make decisions, and resolve conflicts in a team. For example, individuals from countries with low power distance (e.g., Austria, Israel, Ireland, and United States) try to minimize inequalities and favor less autocratic leadership and less centralization of authority in teamwork than do individuals from countries with high power distance (e.g., Malaysia, Philippines, Panama, Guatemala, and Puerto Rico). In addition, countries differ significantly in their expression of collectivism, a difference that is likely to affect an individual's desire to participate in teams (i.e., collective orientation) and the extent to which individuals take advantage of the benefits offered by teamwork.

From the standpoint of developing a measure of teamwork for ALL, the research on culture has two important implications. First, although it seems safe to conclude that the core dimensions of teamwork (see Figure 1) generalize to most countries, it also seems likely that the way in which these skills are manifested will vary by nation. For example, communication will be central to teamwork regardless of culture, but team members from different countries may employ somewhat different communication strategies. In an attempt to address this issue, we tried to identify behaviors representing each of the core teamwork skills that were least likely to vary. However, the extent to which we achieved this goal will only be known after testing the teamwork measure in several different countries. Second, because effective teamwork behaviors likely vary across countries, it may not be possible to construct teamwork items with one "correct" answer. What is considered appropriate team behavior in one country may not be considered appropriate in another. Therefore, our items will attempt to capture information about respondents' knowledge of teamwork across the countries participating in ALL. Norms on these measures will be produced for each country, thereby providing a wealth of information on the nature of teamwork within a country. To the extent that teamwork is manifested differently from culture to culture, cross-cultural comparisons will be neither possible nor appropriate. With these issues in mind, we now turn to a discussion of ALL teamwork measures.

4.3 ALL teamwork measures

The previous sections of this framework have presented our definition of a team and have delineated the core knowledge, skills and attitudes that are associated with effective teamwork (see Figure 1). We have tried to identify individual-level competencies that are generalizable, although we recognize that culture may play a significant role in how individuals express these competencies while functioning in a team.

This section of the framework describes our strategies for assessing teamwork. We first present a series of theoretical and practical assumptions that will guide item development. We present these assumptions here because they have significantly influenced our measurement approach.

4.3.1 Theoretical assumptions

There are four distinguishing features of a team (two or more individuals; a shared or common goal; task interdependence; and a desired productive outcome).

There are generic team competencies held at the individual level that we believe can be measured.

The competencies defined in this framework represent key elements of teamwork that should be measured.

The competencies defined in this framework are critical for successful teamwork.

Attitudes toward teamwork and knowledge of teamwork skills directly affect teamwork.

There are cultural differences associated with teamwork. All cultures will be familiar with the notion of teams, and the competencies reflected in the framework are likely to be common to all cultures. However, these competencies are not necessarily expressed in the same way.

4.3.2 Practical assumptions

Participants will have approximately 30 minutes to complete the Teamwork section of ALL.

Teamwork will be assessed using paper-and-pencil measures.

Although we expect cultural differences in teamwork, we are not trying to measure differences in culture; rather, we emphasize general factors of teamwork with strong cross-cultural relevance.

The same measurement approach will be used to assess teamwork across cultures.

Respondent experience with teams may be work or non-work related (e.g., sports, community, schools, etc.)

Although team processes cannot be directly observed, knowledge about team skills, attitudes toward teamwork, and historical experience with teamwork can be measured.

Among these assumptions, the final practical assumption is most important. As mentioned in the Introduction, it will not be possible to measure respondent team skill competencies directly because the teamwork measure in ALL will be a short paper-and-pencil measure. Measuring team skills has historically required detailed simulations in which team member behaviors are observed and evaluated (D. Baker and Salas, 1992; 1997; Brannick et al., 1997; Ilgen, 1999). Such procedures are inconsistent with the measurement approach of ALL. However, even with these constraints, it is possible to learn a great deal about both the nature of teamwork, and about critical variables that can affect team performance. In particular, respondents' knowledge of teamwork skills (see Figure 1) and respondents' attitudes toward teamwork can be assessed in situation-based items that elicit behavior-oriented, rather than "textbook," responses. Our strategies for measuring each are detailed below.

4.4 Knowledge of teamwork skills

The primary goal of the ALL teamwork measure will be to measure respondent knowledge of teamwork skills, which have been shown to be positively related to team performance (Salas et al., 1999; Stevens and Campion, 1994b). In particular, respondent knowledge of Group Decision Making/Planning, Adaptability/Flexibility, Interpersonal Relations, and Communication will be assessed. Results from this measure will provide information as to how knowledge of teamwork skills is distributed in the adult population within nations.

4.4.1 Measurement approach

In developing our approach for measuring knowledge of teamwork, we faced two significant challenges: (a) because ALL is the first attempt to assess knowledge of teamwork internationally, results from prior research were not available for guidance; and (b) due to practical constraints associated with ALL, the method of measurement was limited to a short paper-and-pencil instrument. Future large-scale assessments of teamwork may consider the use of computer-based simulations or other similar formats to assess team skills more directly (E. Baker, 1998); however, the necessary technology is not currently available to the ALL.

Based on our definition of teamwork, the relevant literature on knowledge tests (Borman, 1991; Dye, Reck and McDaniel, 1993; Hunter, 1986), the domain we sought to measure, and our desire to assess applied knowledge, our questions require respondents to make situational judgments. In personnel selection, both situational judgment questions for written tests and structured interviews have been shown to predict job performance (M. Campion, J. Campion, and Hudson, 1993). Specific to teams, Stevens and Campion (1994b) have reported significant criterion-related validities with supervisory and peer ratings of team performance for a thirty-five-item situational judgment test of teamwork knowledge (although this measure was also significantly correlated with respondent general mental ability). Finally, situational judgment tests have a high degree of face validity for the respondent.

4.4.2 Item development

Initially, an item production grid was constructed to guide item development (refer to Appendix 1.1). The item production grid was derived from the team skill definitions and the behavioral facets representing each skill (i.e., the item production grid in Appendix 1.1 represents the key facets of teamwork in the U.S. and will be modified for different ALL countries). The item production grid is used to ensure that an adequate number of items are developed to cover the skill domains of interest and to specify clearly what each item is intended to measure.

Regarding item construction, short vignettes were initially created. These vignettes describe a fictitious team performing a fictitious team task. Care was taken to ensure that vignettes were based on both work and non-work team situations. Each team described in the vignettes conformed to the definition and characteristics of a "team." To date, five vignettes have been created: one focusing on a toy manufacturing team, one focusing on a marketing team, one focusing on a customer service team and two focusing on community-based teams (one assigned to review school performance and one assigned to clean a park).

Situational judgment items were developed for each vignette. Each item presents a situation, and respondents are asked to rate the effectiveness of each response option on a 5-point scale where 1 indicates "Extremely Bad" and 5 indicates "Extremely Good." To date, eight items have been developed for each vignette, resulting in a total of 40 items. Appendix 1.2 presents several example items. Appendix 1.3 lists all of the items developed thus far.

One issue that was considered, though not specifically accounted for during item development, was the notion of item difficulty. First, unlike other measures included in ALL (i.e., literacy, numeracy, problem solving, etc.), the assessment of teamwork skills (or knowledge of teamwork skills) in the adult population internationally is a new undertaking. Therefore, no research was available to help identify the attributes that might comprise a more difficult and less difficult teamwork item. Certainly, varying the degree to which it is easy to identify the best response from a series of distractors would affect item difficulty. Though this could be done, the ability to respond to more difficult items constructed in this manner would not necessarily reflect more knowledge of teamwork skills. Such responses may be more reflective of a test taker's ability to read, comprehend, and extract the correct information. More importantly, we must acknowledge that the difficulty of teamwork may lie in the execution of team behaviors rather than in the knowledge of what to do. All team members may know what to do in a given team situation, but only the best team members are willing and able to carry out these behaviors in a timely and appropriate fashion that maximizes teamwork. The paper-and-pencil measurement approach used in ALL does not allow for assessing a respondent's skills in terms of actual outcome criteria.

With these issues in mind, we tried to construct items of moderate difficulty. Psychometrically speaking, items of medium difficulty will provide maximum information on the distribution of knowledge of teamwork skills within each ALL country (Crocker and Algina, 1986). Items of medium difficulty were formulated by embedding the "best" alternative for each situational-judgment item ("best" in terms of U.S. research findings) among two alternatives that might reflect other cultures' expressions of team skills and one distracter that virtually no one would be expected to select. In addition, we plan to collect sufficient data during pre-feasibility and feasibility studies to determine each item's difficulty statistically. Item difficulty and other indicators of item performance will be used to select final items for the ALL teamwork measure.

4.4.3 Scoring

Several scoring procedures will be explored during feasibility testing. These range from a Thurstone-like scaling procedure (Anastasi, 1988) in which respondent ratings are compared to country-specific profiles generated for each ALL country to a dichotomous scoring procedure in which each situational judgment item is scored as right or wrong. Whether or not right versus wrong scoring is plausible will be determined by the extent to which rating profiles (i.e., respondent average ratings for the knowledge items) are similar across countries. The final scoring procedure for the team knowledge measure will be selected on the basis of these analyses and practical considerations associated with administration and scoring of ALL.

4.5 Attitudes toward teamwork

Team attitudes are defined as an internal state that influences a team member's choices or decisions to act in a particular way (Cannon-Bowers et al., 1995; Dick and Carey, 1990). Attitudes toward teamwork can have a significant effect on how teamwork skills are actually put into practice. Positive attitudes toward teamwork (Gregorich et al., 1990; Ruffell-Smith, 1979; Helmreich et al.,1986) and an attraction to being part of a team (i.e., collective orientation) have been found to enhance team process and team performance (Driskell and Salas, 1992; Eby and Dobbins, 1997). Therefore, each of these attitudes will be assessed as part of the ALL teamwork measure.

4.5.1 Measurement approach

Unlike the knowledge of teamwork skills, a significant body of work exists on the assessment of attitudes toward teamwork both in the US and internationally (see for example, Eby and Dobbins, 1997; Gregorich et al., 1990; Helmreich et al., 1986). The vast majority of this work, however, has focused on commercial pilot attitudes toward teamwork in the cockpit. Nonetheless, this research provides an excellent starting point for structuring our measurement approach.

A review of past attitude measures indicated that all employed some form of Likert scaling. A similar approach is proposed for ALL. Likert-type scales typically include a series of positive and negative statements about teamwork, and respondents endorse one of a series of graded response options (e.g., strongly agree, agree, neutral, disagree, strongly disagree) for each item. Points are allocated to each response option (e.g., 5 = strongly agree, 4 = agree, etc.) and the sum of these values represent attitude strength.

4.5.2 Item development

Positive and negative statements regarding Belief in the Importance of Teamwork and Collective Orientation were identified and extracted from the research on team attitude measurement (Eby and Dobbins, 1997; Gregorich et al., 1990). Some of these statements were rephrased because they were extracted from a measure designed to assess pilot attitudes toward teamwork in the cockpit. In addition, several new statements were prepared to ensure that a sufficient number of statements were included for reliable measures. In total, 16 statements were developed to measure Belief in the Importance of Teamwork and 15 statements were developed to measure Collective Orientation. Consistent with other approaches, all statements were scaled using a five-point Likert-type scale where 1 = strongly disagree and 5 = strongly agree.

Attitude measures were tested on 192 business students from a mid-western university. Of the cases in which complete demographic data were available, 173 were undergraduate students, 2 were graduate students, and 1 was a professor. The mean age of participants was 22 years old; 74 were female and 108 were male; 73.4% were Caucasian, 12% were Asian or Pacific Islander, 5.2% were African-American and 1% were Hispanic. Most participants (93.8%) also indicated that they had some experience working or participating in a team.

A principal components factor analysis, item-subscale correlations, and a qualitative review of the clarity and potential for cultural bias associated with each item were used to select final items for the two attitude scales. This process resulted in eight items being selected to measure Belief in the Importance of Teamwork (μ = .79) and seven items being selected to measure Collective Orientation (μ = .84). Appendix 1.4 contains the final items for the Team Attitude Scale.

4.5.3 Scoring

The Belief in the Importance of Teamwork Scale and the Collective Orientation Scale will be scored in the same fashion. Total scores will be calculated by summing the points associated with the response alternatives selected by each respondent. Negative statements will be reverse-coded. Scores on the Belief in the Importance of Teamwork Scale can range from a low of 5 to a high of 40, whereas scores on the Collective Orientation Scale can range from a low of 5 to a high of 35. Norms will be developed on the basis of these scale scores, thereby providing information regarding attitudes toward teamwork for countries participating in ALL.

5. Background questionnaire

The Background Questionnaire presents an opportunity to collect information about demographic, social, and economic factors that affect teamwork. Such information should be of interest to policymakers and educators from countries participating in ALL because it will provide information on the determinants of teamwork. Results can be used for structuring policy and/or educational programs to improve the levels of teamwork in the workforce and elsewhere.

Based on what is currently known about teamwork and our approach to measuring teamwork in ALL, we hypothesize that several background variables may have an effect on a respondent's knowledge of team skills, and his or her belief in the importance of teamwork and collective orientation. Specifically, past experience in teams, whether or not the respondent has received formal or informal team training, and demographic variables like respondent age, gender, economic status, and educational level may have an effect. Each of these is briefly discussed in some detail below.

5.1 Experience in teams

The nature and extent of respondents' experiences in teams are likely to significantly affect their attitudes toward teamwork and knowledge of what to do in teams. In addition to including questions about respondents' experiences in teams in the background questionnaire, we developed a short team experience measure (see Appendix 1.5). This scale asks respondents to rate their past experiences in teams on a series of bipolar adjectives. An initial version of this scale was tested on the sample of business students described earlier (refer to item development for the team attitude scales for a detailed description of the sample). Results indicated that the scale was reasonably reliable ($\mu = .79$).

5.2 Team training

Whether or not respondents have received formal or informal team training is likely to have a significant effect on both knowledge of teamwork skills and attitudes toward teamwork. Sufficient research exists to support the efficacy of team training for improving attitudes toward teamwork, increasing knowledge, and enhancing teamwork skills (see for example, Salas et al., 1995; Salas et al., 1999). Collecting information on whether or not ALL respondents have received team training and the nature of training content should prove useful to policymakers interested in improving teamwork skills in the workforce. Data collected through ALL could provide significant insight into specific training strategies that are effective in different countries.

5.3 Demographics

Demographic characteristics such as age, gender, economic status, and educational level may also affect respondent knowledge and attitudes toward teamwork. Little, if any, research on the effects of these variables currently exists. ALL could present an opportunity to assess whether or not knowledge of teamwork skills and attitudes toward teamwork vary as a function of these and other demographic characteristics.

6. Overall summary and conclusions

In closing, this paper presented our framework for assessing teamwork as part of ALL. The framework was derived from the literature on teams and on what is currently known about effective team performance. We have tried to capture the fundamental constructs underlying effective teamwork, in the anticipation that these constructs will generalize to a wide variety of countries, even if their expression may differ across cultures.

In addition to delineating the key facets of teamwork, this paper has also presented our approach to measurement. Although it will not be possible to measure team skill competencies through direct observation, we will nevertheless assess respondents' knowledge of teamwork skills and respondents' attitudes toward working in teams. The results will provide insight into the distribution of these constructs in the international adult population.

Finally, we view this framework as a work in progress for two reasons. First, as with all survey development, we envision conducting significant pilot testing on the approaches we have selected. This testing is likely to lead to revision of our measurement strategies. Although the strategies we have suggested have been effective in other domains, their efficacy for assessing teamwork on an international level has yet to be determined. Second, new information becomes available on teams and the nature of teamwork almost daily. The field is growing and changing concurrently with our efforts. In response, we view our framework as evolving as well; thus, we will incorporate relevant new findings as they become available.

References

Adler, N.J. (1986). *International dimensions of organizational behavior*. Boston, MA: Kent.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan

Ayman, R., and Chemers, M.M. (1983). Relationship of supervisory behavior ratings to work group effectiveness and subordinate satisfaction among Iranian managers. *Journal of Applied Psychology, 68,* 338-341.

Baker, D.P., and Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors*, *34*, 469-475.

Baker, D.P., and Salas, E. (1997). Principles for measuring teamwork: A summary and look toward the future. In M.T. Brannick, E. Salas, and C. Prince (Eds.). *Team performance assessment and measurement* (pp. 331-355). Mahwah, NJ: Erlbaum.

Baker, E.L. (1998). *Draft final report for the validation of teamwork skills questionnaire using computer-based teamwork simulations.* University of California, Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M.D. Dunnette and L. M. Hough (Eds.) *Handbook of industrial and organizational psychology* (pp. 269-313). Palo Alto, CA: Consulting Psychologists Press.

Boyatzis, R.E. (1982). *The competent manager*. New York: John Wiley and Sons.

Brannick, M.T., Salas, E., and Prince, C. (1997). *Team performance assessment and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brooks, A.K. (1994). Power and the production of knowledge: Collective team learning in work organizations. *Human Resource Development Quarterly, 5,* 213-235.

Campion, M.A., Campion, J.E., and Hudson, J.P. (1994). Structured interviewing: A note on incremental validity and alternate question types. *Journal of Applied Psychology, 79,* 998-1002.

Cannon-Bowers, J.A., Tannenbaum, S.I., Salas, E., and Volpe, C.E. (1995). Defining competencies and establishing team training requirements. In R. Guzzo and E. Salas (Eds.), *Team effectiveness and decision making in organizations*, (pp. 333-380). San Francisco: Jossey Bass.

Cannon-Bowers, J.A., and Salas, E. (1997). A framework for developing team performance measures in training (pp. 45-62). In M.T. Brannick, E. Salas, and C. Prince (Eds.). *Team performance assessment and measurement* (pp. 331-355). Mahwah, NJ: Erlbaum.

Carnevale, A.P., Gainer, I.J., and Meltzer, A.S. (1990). *Workplace basics: The essential skills employers want*. San Francisco, CA: Jossey-Bass.

Cohen, S.G., and Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23,* 239-290.

Commission on the Skills of the American Workforce. (1990). *America's choice: High skills or low wages!* Rochester, NY: Author.

Conference Board of Canada (1993). *Employability skills profile*. Ottawa, Ont., Canada: Author [The Conference Board of Canada, 255 Smyth Road, Ottawa, Ont. K1H 8M7, Canada, 10/93]

Crocker, L., and Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Reinhart, and Winston.

Dick, W., and Carey, L. (1990). *The systematic design of instruction* (3rd ed.). Glenview, IL: Scott Foresman.

Driskell, J.E., and Salas, E. (1992). Collective behavior and team performance. *Human Factors*, *34*, 277-288.

Dye, D.A., Reck, M., and McDaniel, M.A. (1993). Moderators of the validity of written job knowledge measures. *International Journal of Selection and Assessment*, *1*, 153-157.

Dyer, J. (1984). Team research and training: A state-of-the-art review. In F.A. Muckler (Ed.), *Human Factors Review: 1984* (pp. 285-323). Santa Monica, CA: Human Factors Society.

Eby, L.T., and Dobbins, G.H. (1997). Collectivistic orientation in teams: An individual and group level analysis. *Journal of Organizational Behavior, 18,* 275-2795.

Gibson, C.G. (1994). The implications of culture for organization structure: An investigation of three perspectives. *Advances in International Comparative Management*, *9*, 3-38.

Gibson, C.B. (1996). *They do what they believe they can? Group-efficacy beliefs and group performance across tasks and cultures*. Paper presented at the National Academy of Management meetings, Cincinnati, OH.

Gibson, C.B., and Zellmer, M.E. (1997). *Intercultural analysis of the meaning of teamwork: Evidence from six multinational corporations*. Paper presented at the Academy of International Business Conference, Monterrey, Mexico.

Gordon, J. (1992). Work teams: How far have they come? *Training*, (October 1992), 59-65.

Gregorich, S.E., Helmreich, R.L., and Wilhelm, J.A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75,* 682-690.

Guzzo, R.A., and Shea, G.P. (1992). Group performance and inter-group relations in organizations. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., 3, 269-313). Palo Alto, CA: Consulting Psychologists Press, 1992.

Hackman, J.R. (1987). The design of work teams. In J.W. Lorsch (Ed.). *Handbook of Organizational Behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice Hall.

Helmreich, R.L., Foushee, H.C., Benson, R., and Russini, W. (1986). Cockpit management attitudes: Exploring the attitude-behavior linkage. *Aviation, Space, and Environmental Medicine, 57,* 1198-1200.

Helmreich, R.L., Merritt, A.C., and Sherman, P.J. (1996). Human factors and national culture. *International Civil Aviation Organization Journal, 51*, 14-16.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.

Hofstede, G. (1985). The interaction between national and organizational value systems, *Journal of Management Studies, 22*, 347-357.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind.* UK: McGraw-Hill.

Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

Ilgen, D.R. (1999). Teams embedded in organizations. *American Psychologist, 54,* 129-139.

Kirkman, B.L. (1997). The impact of cul-tural values on employee resistance to teams: Toward a model of globalized self-managing work team effectiveness. A*cad-emy of Management Review, 22*, 730-757.

Leslie. J., and Van Velsor, E. (1998). *A Cross-National Comparison of Effective Leadership and Teamwork.* Greensboro, NC: Center for Creative Leadership.

Macionis, J.J. (1997*). Sociology* (6th ed.) Upper Saddle River, NJ: Prentice Hall.

McGrath, J.E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice Hall.

McGrath, J. (1997). Small group research, that once and future field: An interpretation of the past with an eye to the future. *Group Dynamics: Theory, Research, and Practice, 1(1),* 7-27.

McIntyre, R.M., and Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. Guzzo, E. Salas, and Associates (Eds.), *Team effectiveness and decision making in organizations,* (pp. 9-45). San Francisco, CA: Jossey-Bass.

Merritt, A.C. (1996). *National culture and work attitudes in commercial aviation: A cross-cultural investigation.* Unpublished doctoral dissertation. The University of Texas at Austin.

Merritt A.C., and Helmreich, R.L. (1996-a). Human factors on the flight deck: The influence of national culture. *Journal of Cross-Cultural Psychology, 27,* 5-24.

Morgan, B.B., Glickman, A.S., Woodward, E.A., Blaiwes, A., and Salas, E. (1986). *Measurement of team behaviors in a Navy environment* (NTSC Report No. 86-014). Orlando, FL: Naval Training Systems Center.

O'Neil, H.F., Chung, G.K.W.K., Brown, R.S. (1997). Use of network simulations as a context to measure team competencies. In H.F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Oser, R.L., McCallum, G.A., Salas, E., and Morgan, B.B., Jr. (1992). Toward a definition of teamwork: Behavioral elements of successful teams (*NTSC Technical Report No. 89-018*). Orlando, FL: Naval Training Systems Center.

Parry, S.B. (1998). Just what is a competency (And why should we care?). *Training*, 58-64.

Pizam, A., and Reichel, A. (1977). Cultural determinants of managerial behavior. *Management International Review, 17*, 65-72.

Ruffell-Smith, H.P. (1979). A simulator study of the interaction of pilot workload with errors (NASA Technical Report No. TM-78482). Moffett Field, CA: National Aeronautics and Space Administration-Ames Research Center.

Salas, E., Bowers, C.A., and Cannon-Bowers, J.A. (1995). Military team research: 10 years of progress. *Military Psychology, 7*, 55-75.

Salas, E., Dickinson, T., Converse, S.A., and Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In R.W. Swezey and E. Salas (Eds.), *Teams: Their training and performance*, (pp. 3-29). Norwood, NJ: Ablex.

Salas, E., Fowlkes, J.E., Stout, R.J., Milonvich, D.M., and Prince, C. (1999). *Does CRM training improve teamwork skills in the cockpit? Two evaluation studies*. Submitted for publication.

Stevens, M.J., and Campion, M.A. (1994a). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management, 20*, 503-530.

Stevens, M.J., and Campion, M.A. (1994b). Staffing teams: Development and validation of the teamwork KSA test. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Nashville, TN.

Triandis, H.C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review, 96*, 506-520.

Trompenaars, F. (1994). *Riding the waves of culture: Understanding diversity in global business*. Chicago, IL: Irwin Professional Publishing.

US Department of Labor (1991, June). *What work requires of schools: A SCANS report for American 2000*. Washington, DC: US Department of Labor, The Secretary's Commission of Achieving Necessary Skills (SCANS).

US Department of Labor (1992a). *Learning and living: A blueprint for high performance*. Washington, DC: US Department of Labor, The Secretary's Commission of Achieving Necessary Skills (SCANS).

US Department of Labor (1992b). *Skills and tasks for jobs: A SCANS report for American 2000*. Washington, DC: US Department of Labor, The Secretary's Commission of Achieving Necessary Skills (SCANS).

Vaziri, M.T., Lee, J.W., and Krieger, J.L. (1988). Onda Moku: The true pioneer of management through respect for humanity. *Leadership and Organization Development Journal, 9*, 3-7.

Appendices

Appendix 1.1

Item Production Grid

Teamwork Skill	Behavioral Requirements	Items
Group Decision Making/Planning	Identify problems	
	Gather information	
	Evaluate information	
	Share information	
	Understand decisions	
	Set goals	
Adaptability/Flexibility	Provide assistance	
	Reallocate tasks	
	Provide/Accept feedback	
	Monitor/Adjust performance	
Interpersonal Relations	Share the work	
	Seek mutually agreeable solutions	
	Consider different ways of doing things	
	Manage/Influence disputes	
Communication	Provide clear and accurate information	
	Listen effectively	
	Ask questions	
	Acknowledge requests for information	
	Openly share ideas	
	Pay attention to non-verbal behavior	

Appendix 1.2

Example Knowledge Items

The following survey describes a team and several situations that the team encounters. After each situation, there are several response options describing what the team could do. For each option listed, rate the quality of the option on the following 1-to-5 scale.

Rating Scale

1	2	3	4	5
Extremely Bad	Somewhat Bad	Neither Bad Nor Good	Somewhat Good	Extremely Good

Vignette 4

A team of volunteers cleans a community park each month. The park is so large that the team needs an entire day to clean it.

Item 1

Members of the team have always worked well together. Recently, the community requested that the park be cleaned more often. The team meets to discuss this requirement, but team members disagree about how to proceed. To help this situation, team members should:

a) _____ Act as though the differences in opinion are not very important.

b) _____ Write down the various opinions about how to proceed and have a team member select one at random.

c) _____ Ask someone from outside the team to act as a mediator at the next meeting.

d) _____ Conduct a candid discussion about the issues on which the team members disagree.

Item 2

The team is asked to periodically rake all the leaves in the park every few weeks during the fall. This situation places a new demand on the team. To cope with this increased demand on its time, the team should:

a) _____ Refuse to do the additional work.

b) _____ Distribute the additional work equally among team members.

c) _____ Assign the additional work to the newest team member.

d) _____ Ask another team to do half the work.

Item 3

One team member leaves the team and a new individual volunteers. The next month the park is cleaned, the team should:

a) _____ Assign the new team member his fair share of the work, but be willing to help out, if necessary.

b) _____ Assign the new team member only the easiest tasks.

c) _____ Encourage the new team member to learn the work by trial and error.

d) _____ Tell the new team member to stay out of the way and watch what the other team members are doing.

Item 4

No one on the team wants to clean the park bathrooms. To resolve this situation, the team should:

a) _____ Decide through a lottery who cleans the bathrooms each time.

b) _____ Have the newest team member clean the bathrooms.

c) _____ Rotate the responsibility of cleaning the bathrooms to a different team member each month.

d) _____ Refuse to clean the bathrooms, since no one on the team wants to do it.

Item 5

The team is requested to make a recommendation on how to improve the park. When the team meets to decide on its recommendation, the team should:

a) _____ Discuss a wide variety of recommendations before making a decision.

b) _____ Allow each team member to suggest one recommendation for consideration by the team.

c) _____ Assign the responsibility for making a recommendation to the team member who seems to know the most about parks.

d) _____ Tell the community it is not the team's job to make a recommendation.

Item 6

The next park cleaning is scheduled for a holiday and most team members will be out of town. The team meets to reschedule cleaning the park. During this meeting, team members should:

a) _____ Try to participate as much as possible in the decision making process.

b) _____ Hide their own feelings to promote good relationships.

c) _____ Anticipate and discuss potential problems with cleaning the park on a different day.

d) _____ Encourage quieter team members to go along with the most outspoken members in order to reach a quick decision.

Item 7

While cleaning the park, a team member is uncertain about what another team member has asked him to do. The team member should:

a) _____ Try to guess what the other team member wanted.

b) _____ Ignore the request; the other team member will ask again if it's important.

c) _____ Ask the other team member to repeat what he or she said.

d) _____ Tell the other team member to speak more clearly.

Appendix 1.3

Teamwork Situational Judgment Items

Teamwork Skill	Behavioral Requirements	Items
Group Decision Making/ Planning	Identify problems	V4-I6
	Gather information	V3-I1
	Evaluate information	V1-I2; V1-I8; V2-I6; V3-I2
	Share information	V4-I5; V5-I1
	Understand decisions	V3-I6; V5-I5
	Set goals	V2-I2; V5-I2
Adaptability/Flexibility	Provide assistance	V1-I1; V4-I3
	Reallocate tasks	V2-I4; V4-I2
	Provide/Accept feedback	V1-I7; V3-I4; V5-I4
	Monitor/Adjust performance	V2-I5; V3-I3; V5-I3
Interpersonal Relations	Share the work	V4-I4
	Seek mutually agreeable solutions	V1-I3; V1-I4
	Consider different ways of doing things	V2-I1; V5-I6
	Manage/Influence disputes	V2-I3; V3-I5; V4-I1
Communication	Provide clear and accurate information	V3-I8
	Listen effectively	V2-I6; V4-I8
	Ask questions	V4-I7; V5-I7
	Acknowledge requests for information	V1-I5
	Openly share ideas	V2-I7; V2-I8
	Pay attention to non-verbal behavior	V3-I7; V5-I8

Note: V2-I8 indicates Vignette 2 — Item 8.

Appendix 1.4

Team Attitude Scale

For each item, please indicate your response by circling the appropriate number for each item in the scale below.

	Strongly disagree		Neither agree or disagree		Strongly agree
1. Teamwork skills deserve more attention in the workplace.	1	2	3	4	5
2. Teams make better decisions than individuals.	1	2	3	4	5
3. Given a choice, I would rather work alone than do a job where I have to work in a team.	1	2	3	4	5
4. It is impossible to function in today's society without being a good team player.	1	2	3	4	5
5. I prefer to participate in team-oriented activities.	1	2	3	4	5
6. Teams always outperform individuals.	1	2	3	4	5
7. Everyone should be taught to be a good team player.	1	2	3	4	5
6. Teams always outperform individuals.	1	2	3	4	5
7. Everyone should be taught to be a good team player.	1	2	3	4	5
8. I prefer to work on teams where team members perform their own tasks independently rather than working together.	1	2	3	4	5
9. I find that working as a member of a team increases my ability to perform effectively.	1	2	3	4	5
10. I find working in a team to be very satisfying.	1	2	3	4	5
11. Teamwork is one of the most important skills in life.	1	2	3	4	5
12. I prefer to be rewarded for my team's performance rather than my individual performance.	1	2	3	4	5
13. People with strong teamwork skills will always be successful.	1	2	3	4	5
14. Teams plan better than individuals.	1	2	3	4	5
15. I prefer working as part of a team to working alone.	1	2	3	4	5

Appendix 1.5

Team Experience

For each word pair, please assess your overall past experience across all of the teams you have participated in by circling the appropriate number on the scale provided.

Note: If you have never worked/participated in a team, please provide your perceptions as to what you think working/participating in most teams would be like.

Example Items

		Neutral
1. Competitive	1 ——— 2 ——— 3 ——— 4 ——— 5 ——— 6 ——— 7	Cooperative
		Neutral
2. Open	1 ——— 2 ——— 3 ——— 4 ——— 5 ——— 6 ——— 7	Closed
		Neutral
3. Rigid	1 ——— 2 ——— 3 ——— 4 ——— 5 ——— 6 ——— 7	Flexible
		Neutral
4. Trusting	1 ——— 2 ——— 3 ——— 4 ——— 5 ——— 6 ——— 7	Distrustful
		Neutral
5. United	1 ——— 2 ——— 3 ——— 4 ——— 5 ——— 6 ——— 7	Divided

Appendix 2.1

Frequencies for Canada and Italy for the teamwork modules during the feasibility studies

Section L Teamwork - Past Experience – Frequencies – Canada only

Below you will find a list of various types of teams along with a description of each type and some examples.

Please read the descriptions and examples and then indicate the context in which you have ever participated in any of the types of teams. Then, tell us how much experience you have had in each of the teams in which you participated.

Type of Team and Description	Did you participate in this type of team in a work organization?	Did you participate in this type of team in a student organization?	Did you participate in this type of team in a volunteer or community organization?	Did you participate in this type of team in a religious organization?
L11	L1BA L1AA	L1D L1CA	L1F L1EA	L1HA L1G
Management team Team is responsible for coordinating, budgeting, and staffing several units within an organization. **Examples:** Executive and Management teams	Yes 1 30 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 70 ➜ **Go to question L1C**	Yes 1 15 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 84 ➜ **Go to question L1E**	Yes 1 29 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 70 ➜ **Go to question L1G**	Yes 1 8 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 91 ➜ **Go to question L2 next type of team**
L2	L2BA L2AA	L2DA L2CA	L2F L2EA	L2HA L2G
Project Teams (task forces) Team is responsible for accomplishing a specific set of tasks (develop and carry out an idea, plan, etc.) within a given period of time. The team then disbands when the task is complete. **Examples:** New-product teams, design teams	Yes 1 36 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 64 ➜ **Go to question L2C**	Yes 1 25 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 74 ➜ **Go to question L2E**	Yes 1 31 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 69 ➜ **Go to question L2G**	Yes 1 7 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 92 ➜ **Go to question L3 next type of team**

Appendix 2.1 - continued

Frequencies for Canada and Italy for the teamwork modules during the feasibility studies

Below you will find a list of various types of teams along with a description of each type and some examples.

Please read the descriptions and examples and then indicate the context in which you have ever participated in any of the types of teams. Then, tell us how much experience you have had in each of the teams in which you participated.

Type of Team and Description	Did you participate in this type of team in a work organization?		Did you participate in this type of team in a student organization?		Did you participate in this type of team in a volunteer or community organization?		Did you participate in this type of team in a religious organization?	
L31	L3BA L3AA		L3DBA L3CA		L3FA L3EAA		L3HBA L3G	
Production teams Team is responsible for repeatedly producing a specific product. **Examples:** assembly teams, coal-mining crews	Yes [1] 22 ➜ No [2] 79 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L3C**	Yes [1] 5 ➜ No [2] 94 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L3E**	Yes [1] 9 ➜ No [2] 90 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L3G**	Yes [1] 5 ➜ No [2] 94 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L4 next type of team**
L41	L4BA L4AA		L4DA L4CA		L4FA L4EA		L4H L4G	
Service Teams Team is responsible for repeatedly providing a specific service. Repeated transactions with customers or clients. **Examples:** Retail sales teams, maintenance teams, airline attendent crews, customer service teams	Yes [1] 53 ➜ No [2] 48 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L4C**	Yes [1] 7 ➜ No [2] 93 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L4E**	Yes [1] 17 ➜ No [2] 84 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L4G**	Yes [1] 3 ➜ No [2] 97 ➜	How much experience have you had in this type of team? [1] ◯ less than 1 year [2] ◯ 1 year [3] ◯ 2 - 3 years [4] ◯ 4 - 5 years [5] ◯ more than 5 years **Go to question L5 next type of team**

Appendix 2.1 - continued

Frequencies for Canada and Italy for the teamwork modules during the feasibility studies

Below you will find a list of various types of teams along with a description of each type and some examples.

Please read the descriptions and examples and then indicate the context in which you have ever participated in any of the types of teams. Then, tell us how much experience you have had in each of the teams in which you participated.

Type of Team and Description	Did you participate in this type of team in a work organization?	Did you participate in this type of team in a student organization?	Did you participate in this type of team in a volunteer or community organization?	Did you participate in this type of team in a religious organization?
L5	**L5BA** **L5AA**	**L5DA** **L5CA**	**L5FA** **L5EA**	**L5H** **L5GA**
Action and Performing Teams Team is responsible for performing a specific event within a given period of time. **Examples:** Patient care teams, cockpit crews, firefighting teams, rescue teams, bands or musician ensembles, performing arts ensembles	Yes 1 21 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 81 ➜ **Go to question L5C**	Yes 1 12 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 88 ➜ **Go to question L5E**	Yes 1 21 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 80 ➜ **Go to question L5G**	Yes 1 3 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 95 ➜ **Go to question L6 next type of team**
L6	**L6BA** **L6AA**	**L6DA** **L6CA**	**L6FA** **L6EA**	**L6H** **L6GA**
Parallel Teams Brought together to generate ideas, make suggestions, recommendations, or to solve a specific problem. **Examples:** Ad hoc committees. Quality Circles, TQM	Yes 1 29 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 69 ➜ **Go to question L6C**	Yes 1 16 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 83 ➜ **Go to question L6E**	Yes 1 14 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 84 ➜ **Go to question L6G**	Yes 1 3 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 95 ➜ **Go to question L7 next type of team**

Appendix 2.1 -cconcluded

Frequencies for Canada and Italy for the teamwork modules during the feasibility studies

Below you will find a list of various types of teams along with a description of each type and some examples.

Please read the descriptions and examples and then indicate the context in which you have ever participated in any of the types of teams. Then, tell us how much experience you have had in each of the teams in which you participated.

Type of Team and Description	Did you participate in this type of team in a work organization?	Did you participate in this type of team in a student organization?	Did you participate in this type of team in a volunteer or community organization?	Did you participate in this type of team in a religious organization?
L77	L7BA L7AA	L7DA L7CA	L7FA L7EA	L7H L7GA
Other Teams **Examples:** Boy or Girl Scouts, 4-H clubs, future Farmers of America, academic clubs (e.g., school newspaper, yearbook, science club), school spirit or fund raising clubs, student council/student government	Yes 1 21 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 79	Yes 1 30 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 70	Yes 1 36 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 63	Yes 1 5 ➜ How much experience have you had in this type of team? 1 ◯ less than 1 year 2 ◯ 1 year 3 ◯ 2 - 3 years 4 ◯ 4 - 5 years 5 ◯ more than 5 years No 2 92

In the last 12 months, were you involved in any team sports (e.g., soccer, basketball, etc.)?

L8A

1 F Yes ➜ **Go to question L8A**
2 F No ➜ **Go to question L9**

Did you ... (MARK ALL THAT APPLY).

L8A

		Yes	No
1 F	Play in an organized league with other teams	16	8
2 F	Play on an informal basis with friends and colleagues	3	8
3 F	Serve as a coach in an organized youth/adult league	0	0
4 F	Serveas a referee in an organized youth/adult league	0	0

How frequently did you play, coach or referee?

L8B

1 12 Two or more times per week
2 7 Once per week
3 0 Once per month
4 0 Two to three times per year

In the last 12 months, how would you describe the extent of your involvement in volunteer, community and/or religious organizations? (MARK ALL THAT APPLY).

L9

1 7 active leader (for example, serve as an officer)
2 19 active member (for example, attends all club meetings and always helps out when needed)
3 23 member - I do some work on occasion
4 5 member - but I=m not really active
5 47 Do not participate in such organizations

In your current job, approximately what percentage of your time at work do you spend working in a team? ***(If you have more than one job, tell us about the one at which you work the most hours.)***

L10

1 20 I do not work on a team
2 9 Less than 20%
3 5 21 B 40%
4 4 41 B 60%
5 8 61 B 80%
6 17 More than 80%
N 39 I do not have a job

In your current job, how important is teamwork for accomplishing your tasks?

L11

1 12 I do not work on a team
2 2 Not Important
3 12 Somewhat Important
4 7 Important
5 18 Very Important
6 13 Extremely important
N 37 I do not have a job

For the following statements below, mark all that apply.

L12

1 F In the last 12 months, I have taken a formal course(s) on teamwork from an educational institution (school, college, university, institute, etc.)
2 F In the last 12 months, I participated in a training program(s) on teamwork provided by my employer
3 F In the last 12 months, I have received on-the-job training on teamwork from my employer.
4 F In the last 12 months, I have read a book on teamwork.

In the last 12 months, please indicate the types of teams (either at work or outside of work) you have been a member of (mark all that apply).

L130

1 F I have not been a member of a team during the last 12 months.
2 F I have worked on a team where team members reported to or took direction from a higher-ranking individual on the team.
3 F I have worked on a team where all members were basically the same rank and one person was elected or appointed as the leader.
4 F I have worked on a team where all members were basically the same rank and we shared responsibilities for coordinating activities.
5 F I have worked on a team where I was dependent upon other team members doing their job in order for me to do mine.
6 F I have worked on a team where I could only meet my goals if other team members met theirs.

M1

Organizations (both work and non-work) throughout the world are increasingly relying on work teams. A work team is defined as a group of individuals who produce goods or services for which they are all accountable. A distinguishing feature of all work teams is that members of the team are "interdependent", that is, members of the team must work together to perform work tasks. No team member can accomplish the team's tasks alone. Below is a series of statements about work teams that are associated with specific things that teams do, such as making decisions, communicating, interacting, etc. Please indicate the extent to which you agree or disagree with each statement.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1) Teams should always choose the first suggestion offered when trying to make an important decision.	1 92	2 38	3 16	4 6	5 3
2) Teams should gather information from a wide variety of sources when making an important decision.	1 0	2 0	3 2	4 48	5 102
3) Team members should try to anticipate potential problems with the team's decision.	1 0	2 1	3 15	4 76	5 59
4) Teams should let the most skilled team member make important decisions for the rest of the team.	1 34	2 60	3 30	4 18	5 8
5) Team members should ignore most team decisions.	1 100	2 42	3 3	4 3	5 1
6) Team members should discuss potential problems with the team's decision.	1 1	2 1	3 9	4 68	5 70
7) Teams should assign one person from the team to make the decision for the rest of the team.	1 76	2 51	3 12	4 9	5 1
8) Teams should make important decisions after evaluating different alternatives.	1 1	2 1	3 6	4 73	5 68
9) Team members should quit the team if they don't like a decision.	1 43	2 56	3 36	4 9	5 3
10) Team members should refrain from voicing their opinions about team decisions.	1 62	2 52	3 17	4 14	5 3
11) Team members should try to understand the reasons for the team's decision.	1 1	2 12	3 14	4 84	5 49
12) Teams should ask someone who is not a member of the team to make an important decision.	1 63	2 54	3 25	4 5	5 2
13) Teams should gather information mainly from the team's supervisor when making a decision.	1 32	2 57	3 33	4 20	5 5
14) Team members should share information when making an important decision.	1 1	2 1	3 2	4 55	5 90
15) Teams members should ignore the schedule and perform the work at their own pace.	1 42	2 66	3 27	4 10	5 3
16) Team members should help other members with the work if they need it.	1 0	2 2	3 6	4 84	5 57
17) Teams should try to get everyone on the team to work at the same pace.	1 6	2 37	3 37	4 53	5 15
18) Teams should revise deadlines for team members who are behind schedule.	1 6	2 46	3 42	4 46	5 8
19) Teams should tell new members to stay out of the way.	1 75	2 63	3 6	4 2	5 2
20) Teams should ask individuals outside of the team how the team is doing.	1 12	2 25	3 44	4 57	5 9
21) Teams should make the person who shows up last do any additional work.	1 58	2 65	3 19	4 4	5 2
22) Team members should make suggestions to other members as to how to improve their performance.	1 4	2 4	3 23	4 81	5 36
23) Teams should assign new team members the easy tasks.	1 18	2 53	3 40	4 34	5 3
24) Teams should distribute new work equally among team members.	1 1	2 4	3 17	4 82	5 44
25) Team members should evaluate each other's performance.	1 5	2 26	3 38	4 61	5 17
26) Teams should punish members who make mistakes.	1 60	2 60	3 18	4 8	5 1
27) Team members should act as though differences of opinion are not very important.	1 30	2 47	3 17	4 43	5 12
28) Team members should be open to different ways of doing things.	1 1	2 0	3 3	4 76	5 68
29) Teams members should be open to suggestions.	1 0	2 2	3 2	4 82	5 63
30) Teams should discourage team members from bringing up differences of opinion.	1 44	2 77	3 14	4 8	5 3
31) Team members should distract other members during team meetings.	1 87	2 49	3 6	4 1	5 2

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
32) Teams should ask someone outside the team to act as a mediator at team meetings.	1 23	2 45	3 53	4 24	5 2
33) Team members should discuss other team members' behaviour without them present.	1 64	2 55	3 18	4 8	5 1
34) Team members should treat other team members with dignity and respect.	1 0	2 1	3 1	4 40	5 105
35) Team members should conduct candid discussions about issues when they disagree.	1 3	2 6	3 16	4 67	5 56
36) Team members should express anger toward members who disagree with the rest of the team.	1 78	2 56	3 8	4 3	5 1
37) Team members should try to find common ground when they disagree.	1 3	2 5	3 9	4 82	5 48
38) Team members should make jokes about other members' behaviour.	1 58	2 51	3 27	4 9	5 1
39) Teams should discipline team members who disagree with the rest of the team.	1 56	2 51	3 12	4 21	5 4
40) Team members should only let the most experienced team members talk.	1 71	2 58	3 12	4 5	5 0
41) Team members should pay attention to other team members' tone of voice.	1 3	2 24	3 40	4 66	5 12
42) Team members should ask questions of other team members.	1 4	2 7	3 15	4 78	5 41
43) Teams members should try to guess what other team members are going to say.	1 40	2 56	3 42	4 6	5 2
44) Team members should ignore other member's requests.	1 73	2 60	3 4	4 6	5 2
45) Team members should always speak in a specific order.	1 23	2 57	3 33	4 23	5 8
46) Team members should only share information that team members must know to do their jobs.	1 24	2 66	3 23	4 26	5 8
47) Team members should provide clear and accurate information to one another.	1 0	2 1	3 6	4 76	5 63
48) Team members should openly share ideas, opinions, and problems.	1 3	2 1	3 6	4 76	5 63
49) Team members should ignore each other's suggestions.	1 80	2 55	3 7	4 3	5 1
50) Team members should respond to other team members' questions.	1 3	2 3	3 8	4 90	5 42
51) Team members should not ask other team members to repeat what they said.	1 38	2 69	3 26	4 10	5 1
52) Team members should spend time talking about activities outside of work.	1 15	2 19	3 60	4 41	5 10

→ **PLEASE GO TO SECTION N**

N1

Below is a series of statements about team attitudes. Please indicate the extent to which you agree or disagree with each statement.

		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1)	Teamwork deserves more attention in the workplace.	1 2	2 8	3 22	4 79	5 37
2)	Teams make better decisions than individuals.	1 5	2 13	3 42	4 64	5 25
3)	Given a choice, I would rather work alone than do a job where I have to work in a team.	1 17	2 52	3 41	4 23	5 14
4)	It is impossible to function in today's world without being a good team player.	1 5	2 16	3 39	4 66	5 24
5)	I prefer to participate in team-oriented activities.	1 2	2 23	3 49	4 56	5 18
6)	Teams always outperform individuals.	1 16	2 42	3 60	4 18	5 12
7)	Everyone should be taught to be a good team player.	1 0	2 8	3 27	4 79	5 35
8)	I prefer to work on teams where team members perform their own tasks independently rather than working together.	1 12	2 33	3 54	4 37	5 11
9)	I find that working as a member of a team increases my ability to perform effectively.	1 2	2 14	3 23	4 87	5 20
10)	I find working in a productive team to be very satisfying.	1 0	2 4	3 23	4 83	5 37
11)	Teamwork is one of the most important skills in life.	1 1	2 13	3 30	4 73	5 31
12)	Teams should ask someone who is not a member of the team to make an important decision.	1 63	2 54	3 25	4 5	5 2
13)	Teams should gather information mainly from the team's supervisor when making a decision.	1 32	2 57	3 33	4 20	5 5
14)	Team members should share information when making an important decision.	1 1	2 1	3 2	4 55	5 90
15)	Teams members should ignore the schedule and perform the work at their own pace.	1 42	2 66	3 27	4 10	5 3

→ **PLEASE GO TO SECTION O**

Appendix 2.2

Results of feasibility study on Team Attitudes Measure

Table 1

Correlations, descriptive statistics, and reliability data for team attitudes scale (combined sample)

Subscale	Belief in teams	Collective orientation	Mean	Standard deviation	N
Belief in teams	1.00		28.45	4.94	141
Collective orientation	0.67	1.00	23.96	4.57	142
Alpha	0.82	0.81			

Note: Alpha for the 15-item measure was 0.88.

Table 2

Correlations, descriptive statistics, and reliability data for team attitudes scale (Canadian sample)

Subscale	Belief in teams	Collective orientation	Mean	Standard deviation	N
Belief in teams	1.00		28.64	4.92	99
Collective orientation	0.61	1.00	24.44	4.65	99
Alpha	0.81	0.82			

Note: Alpha for the 15-item measure was 0.88.

Table 3

Correlations, descriptive statistics, and reliability data for team attitudes scale (Italian sample)

Subscale	Belief in teams	Collective orientation	Mean	Standard deviation	N
Belief in teams	1.00		28.00	5.01	42
Collective orientation	0.67	1.00	22.84	4.20	43
Alpha	0.84	0.76			

Note: Alpha for the 15-item measure was 0.89.

Team Non-Cognitive Measure

Results

Note: The results below are based on the 24 “best” items that we identified from the pool of 52 items that were administered.

Table 4

Correlations, descriptive statistics, and reliability data for non-cognitive measure (Combined sample)

Scale	DM	AF	IR	CM	Mean	Standard deviation	N
Decision making	1.00				26.15	2.79	144
Adaptability/Flexibility	0.48	1.00			24.49	2.92	147
Interpersonal relations	0.54	0.52	1.00		26.20	2.91	143
Communication	0.56	0.52	0.67	1.00	26.08	3.08	142

Note: Alpha for the 24-item measure was 0.88.

Table 5

Correlations, descriptive statistics, and reliability data for non-cognitive measure (Canadian sample)

Scale	DM	AF	IR	CM	Mean	Standard deviation	N
Decision making	1.00				26.61	2.79	101
Adaptability/Flexibility	0.48	1.00			24.80	2.99	101
Interpersonal relations	0.51	0.50	1.00		26.35	3.01	98
Communication	0.52	0.51	0.66	1.00	26.70	2.89	99

Table 6

Correlations, descriptive statistics, and reliability data for non-cognitive measure (Italian sample)

Scale	DM	AF	IR	CM	Mean	Standard deviation	N
Decision making	1.00				25.07	2.52	144
Adaptability/Flexibility	0.40	1.00			23.80	2.65	147
Interpersonal relations	0.63	0.54	1.00		25.89	2.68	143
Comm.	0.52	0.47	0.74	1.00	24.63	3.02	142
Alpha	0.66	0.50	0.68	0.66			

Note: Alpha for the 24-item measure was 0.86.

The following results are reported for the non-cognitive measure and attitude measure as a function of respondent experience in different types of work teams. Data are based on the Canadian sample only. Means denoted by (*) are significantly different at $p<.05$. No significant effects were found for the variables: management teams, parallel teams, action and performing teams, and other teams.

Table 7
Project teams

Scale	Experience	N	Means
Decision making	Yes	36	27.25
	No	63	26.28
Adaptability	Yes	36	25.00
	No	63	24.68
Interpersonal	Yes	34	26.62
	No	62	26.19
Communication	Yes	35	27.63 *
	No	62	26.19 *
Belief in teams	Yes	34	28.76
	No	63	28.60

Table 8
Production teams

Scale	Experience	N	Means
Decision making	Yes	22	27.73 *
	No	78	26.28 *
Adaptability	Yes	22	24.31
	No	78	24.91
Interpersonal	Yes	21	27.19
	No	76	26.12
Communication	Yes	20	27.90 *
	No	78	26.41 *
Belief in teams	Yes	21	28.90
	No	77	28.66

Table 9
Service teams

Scale	Experience	N	Means
Decision making	Yes	53	27.08
	No	47	26.19
Adaptability	Yes	53	25.49 *
	No	47	24.06 *
Interpersonal	Yes	53	27.04 *
	No	44	25.50 *
Communication	Yes	53	27.62 *
	No	45	25.67 *
Belief in teams	Yes	51	28.88
	No	47	28.47
Collective orient.	Yes	52	24.30
	No	46	24.63

Chapter 8

The ALL Practical Cognition[1] Framework

Robert J. Sternberg *Yale University, U.S.*
Elena L. Grigorenko *Yale University, U.S. and Moscow State University, Russia*
Guillermo Gil *Spanish Ministry of Education, U.S.*
Jennifer Hedlund *Yale University, U.S.*

1. This report draws for material on parts of Sternberg, R.J., Forsythe, G.B., Hedlund, J., Horvath, J.A., Wagner, R.K., Williams, W.M., Snook, S., and Grigorenko, E.L. (in press). *Practical intelligence in everyday life.* New York: Cambridge University Press.

Table of Contents

Table of Contents

Table of Contents

Summary

Practical cognition is what most people call common sense. It is the skill needed to adapt to, shape, and select everyday environments. Cognition as conventionally defined may be useful in everyday life, but practical cognition is indispensable. Without some measure of it, one cannot survive in a cultural milieu or even in the natural environment. In our work, we have studied many aspects of practical cognition, although we have concentrated on one particularly important aspect of it—*tacit knowledge*—the procedural knowledge one learns in one's everyday life that usually is not taught and often is not even verbalized. Tacit knowledge includes things like knowing what to say to whom, knowing when to say it, and knowing how to say it for maximum effect. In our work, we have studied tacit knowledge in populations as diverse as business managers, military leaders, university professors, elementary-school teachers, janitors, secretaries, salespeople, and U.S. and rural Kenyan children. Tacit knowledge is so-called because it usually starts off tacit, although over time it can come to be verbalized. It is measured by situational-judgment tests.

Our goal is not to denigrate the importance of more academically-based types of cognition, including literacy, numeracy, academic reasoning and so on. Rather, our argument is that more academic types of cognition are not enough—that successful prediction and, more importantly, understanding of performance in the everyday world requires assessment of practical as well as academic types of cognition.

1. Introduction

Practical (or everyday) cognition is different from the kind of cognition associated with academic success. There are any number of ways in which we see this difference in our everyday lives. We see people who succeed in school and who fail in work, or who fail in school but who succeed in work. We meet people with high academic-test scores who seem inept in their social interactions. And we meet people with low test scores who can get along effectively with practically anyone. Laypersons have long recognized a distinction between academic cognition (book smarts) and practical cognition (street smarts or common sense). This distinction is confirmed by research on the implicit theories of cognition held by both laypersons and researchers (Sternberg, 1985b; Sternberg et al., 1981).

1.1 Academic versus practical cognition

There may be any number of reasons for the apparent difference between academic and practical cognition. We argue that a major source of this difference is the sheer disparity in the nature of the kinds of problems one faces in academic versus practical situations. The problems faced in everyday life often have little relation to the knowledge or skills acquired through formal education or used in classroom activities. Consider the following example of an observation made by Richard Wagner of a garbage collector in Tallahassee, Florida.

> Tallahassee, priding itself on the service it provides to its citizens, requires garbage collectors to retrieve trash containers from the backyards of its residents. Each resident fills a large trash container in his or her backyard rather than placing standard-sized garbage cans on the curbside to be picked up. Trash collectors must locate and retrieve each full container from the backyard, heave it into the truck, and then drag the empty container back to each yard. Many of the garbage collectors are young high school dropouts who, because of their lack of education, might be expected to score poorly on cognition tests. On the surface, the job appears to be more physically than cognitively demanding. Each stop requires two trips to the backyard, one to retrieve the full can, and another to return it when it was empty.
>
> One summer it was noticed that the collection routine had changed after a new, older employee joined the crew. This change involved relaxing the constraint that each household retain the same container. Because the trash bins were issued by the city, and not purchased using personal funds, they were identical. The new routine consisted of wheeling the previous house's empty container to the current house's backyard, leaving it to replace the full can, which was in turn wheeled to the truck to be emptied. Once emptied, this can was wheeled to the backyard of the next house to replace its full can, and so on. The new routine required only one trip to each house, where the previous one required two trips. The new employee's insights cut the work nearly in half. This solution had eluded other garbage collectors and the managers who trained them.

Everyone encounters problems in which solutions are neither readily available nor readily derivable from acquired knowledge. This type of problem solving, frequently experienced in daily life, is referred to as *practical problem solving*. Such problems can be experienced at the work place, or in school, the household, stores, movie theaters, or really, anywhere. There is no consensus on how to define practical problems encountered in life, but building on a distinction made by Neisser (1976), Sternberg and his colleagues (Sternberg, 1985a, 1997a; Wagner and Sternberg, 1986) have classified problems as academic or practical in nature. Academic problems tend to be (a) formulated by others, (b) well-defined, (c) complete in the information they provide, (d) characterized by having only one correct answer, (e) characterized by having only one method of obtaining the correct answer, (f) disembedded from ordinary experience, and (g) of little or no intrinsic interest.

Practical problems, in contrast to academic problems, tend to be (a) unformulated or in need of reformulation, (b) of personal interest, (c) lacking in information necessary for solution, (d) related to everyday experience, (e) poorly defined, (f) characterized by multiple "correct" solutions, each with liabilities as well as assets, and (g) characterized by multiple methods for picking a problem solution. Given the differences in the nature of academic and practical problems, it is no surprise that people who are adept at solving one kind of problem may well not be adept at solving problems of the other kind.

The cognitive skills that individuals exhibit in finding solutions to practical problems may be referred to as *practical cognitive skills* (Baltes, Dittman-Kohli, and Dixon, 1984; Berg, in press; Berg and Sternberg, 1985; Rogoff, 1982; Sternberg, 1985a, 1997a; Wagner, in press). When combined, these skills can be referred to as *practical cognition*, which is defined as cognitive skills that serve to find a more optimal fit between the individual and the demands of the individual's environment, whether by adapting to the environment, changing (or shaping) the environment, or selecting a different environment (Sternberg, 1985a; Sternberg, 1997a). The concept of practical cognition takes into account the distinction presented above between academic and practical tasks. The skills emphasized in formal schooling have limited value if they cannot be used to address practical, everyday problems.

1.1.1 Research on practical problem-solving skill

The research on practical cognition is becoming more and more central to mainstream psychology (see Berg and Klaczynski, 1996, for a review). Initially, the examination of practical cognition issued from a concern that the cognition of adults functioning largely outside the academic environment from the moment they obtained their academic degrees and virtually for the rest of their lives, was evaluated primarily by traditional tests of cognition constructed to predict academic success.

Various aspects of the meaning of the concept of practical cognition are expressed in a number of diverse constructs. Some researchers define everyday cognition as a specific expression of conventional skills that permit adaptive behavior within a distinct class of everyday-life situations (e.g., Willis and Schaie, 1986), whereas others stress the unique nature of practical skills (e.g., Neisser, 1976; Wagner, 1987). Most psychological studies of practical skills focus on solving problems that are ill-structured in their goals and solutions and are frequently encountered in daily life (at home, work, and in dealing with people) (e.g., Cornelius and Caspi, 1987; Denney, 1989).

A number of studies have addressed the relation between practical and academic cognition. These studies have been carried out in a wide range of settings, using a variety of tasks, and with diverse populations. We review some examples of research on problem solving and reasoning. For other reviews see Ceci and Roazzi (1994), Rogoff and Lave (1984), Scribner and Cole (1981), Sternberg and Wagner (1986, 1994), Voss, Perkins, and Segal (1991), and Wagner (in press). Taken together, these studies show that skill measured in one setting (e.g., school) does not necessarily transfer to another setting (e.g., real-world task).

Several studies compared performance on mathematical types of problems across different contexts. Scribner (1984, 1986) studied the strategies used by milk processing plant workers to fill orders. Workers who assemble orders for cases of various quantities (e.g., gallons, quarts, or pints) and products (e.g., whole milk, two percent milk, or buttermilk) are called assemblers. Rather than employing typical mathematical algorithms learned in the classroom, Scribner found that experienced assemblers used complex strategies for combining partially filled cases in a manner that minimized the number of moves required to complete an order. Although the assemblers were the least educated workers in the plant, they were able to calculate in their heads quantities expressed in different base number systems, and they routinely outperformed the more highly educated white collar workers who substituted when assemblers were absent. Scribner found that the order-filling performance of the assemblers was unrelated to measures of school performance, including cognitive test scores, arithmetic test scores, and grades.

Another series of studies of everyday mathematics involved shoppers in California grocery stores who sought to buy at the cheapest cost when the same products were available in different-sized containers (Lave, Murtaugh, and de la Roche, 1984; Murtaugh, 1985). (These studies were performed before cost per unit quantity information was routinely posted). For example, oatmeal may come in two sizes, 10 ounces for \$.98 for 24 ounces for \$2.29. One might adopt the strategy of always buying the largest size, assuming that the larger size is always the most economical. However, the researchers (and savvy shoppers) learned that the larger size did not represent the least cost per unit quantity for about a third of the items purchased. The findings of these studies were that effective shoppers used mental shortcuts to get an easily obtained answer, accurate enough to determine which size to buy. A common strategy, for example, was mentally to change the size and price of an item to make it more comparable with the other size available. For example, one might mentally double the smaller size, thereby comparing 20 ounces at \$1.96 versus 24 ounces at \$2.29. The difference of 4 ounces for about 35 cents, or about 9 cents per ounce, seems to favor the 24-ounce size, given that the smaller size of 10 ounces for \$.98 is about 10 cents per ounce. These mathematical shortcuts yield approximations that are as useful as the actual values of 9.80 and 9.33 cents per ounce for the smaller and larger sizes, respectively, and are much more easily computed in the absence of a calculator. When the shoppers were given a mental-arithmetic test, no relation was found between test performance and accuracy in picking the best values (Lave et al., 1984; Murtaugh, 1985).

Ceci and colleagues (Ceci and Liker,1986, 1988; see also Ceci and Ruiz, 1991) studied expert racetrack handicappers. Ceci and Liker (1986) found that expert handicappers used a highly complex algorithm for predicting post time odds that involved interactions among seven kinds of information. By applying the complex algorithm, handicappers adjusted times posted for each quarter mile on a previous outing by factors such as whether the horse was attempting to pass other horses, and if so, the speed of the other horses passed and where the attempted passes took place. By adjusting posted times for these factors, a better measure of a horse's speed is obtained. It could be argued that the use of complex interactions to predict a horse's speed would require considerable cognitive skill (at least as it is traditionally measured). However, Ceci and Liker reported that the successful use of these interactions by handicappers was unrelated to their overall cognitive ability.

A subsequent study attempted to relate performance at the racetrack to making stock-market predictions in which the same algorithm was involved. Ceci and Ruiz (1991) asked racetrack handicappers to solve a stock-market-prediction task that was structured similarly to the racetrack problem. After 611 trials on the stock-market task, the handicappers performed no better than chance, and there was no difference in performance as a function of overall cognitive ability. Ceci and Roazzi (1994) attribute this lack of transfer to the low correlation between performance on problems and their isomorphs. "Problem isomorphs" refer to two or more problems that involve the same cognitive processes but that use different terminology or take place in different contexts.

The same principle that applies to adults appears also to apply to children. Carraher, Carraher, and Schliemann (1985) studied Brazilian children who, for economic reasons, often worked as street vendors (see also Nuñes, 1994). Most of these children had very little formal schooling. Carraher et al. compared the performance of these children on mathematical problems that were embedded in a real-life situation (i.e., vending) to problems presented in an academic context (e.g., 2 + 4 = ?). The children correctly solved significantly more questions that related to vending than they did math problems that were academic in nature. When the academic problems were presented as word problems (e.g., If an orange costs 76 cruzeiros and a passion fruit cost 50, how much do the two cost together?), the rate of correct responses was substantially better, but still not as high as when the problems were presented in the context of vending.

This lack of transfer also appears to work in the reverse direction. For example, Perret-Clermont (1980) found that many school children had no problem solving paper-and-pencil arithmetic questions, but could not solve the same type of problem in a different context (e.g., counting bunches of flowers). That is, school children may fail to transfer the academic knowledge to everyday problems.

Roazzi (1987) found similar results when comparing street-vendor children to middle-class school children. He compared the performance of children on a class-inclusion task. To assess the performance of the street-vendor children, the researcher posed as a customer and asked questions about the items to find out if the children understood the relationship among classes and subclasses of food (e.g., mint and strawberry chewing gum as part of the class "chewing gum"). At a later time the same children were given a formal test with the same logical structure, but that was irrelevant to their street-vending jobs. The middle-class children were given the same two tests. Street-vendor children performed significantly better on the class-inclusion task in the natural than in the formal context, whereas middle-class children were more successful on the formal version of the task.

Additional research has shown that the use of complex reasoning strategies does not necessarily correlate with overall cognitive ability. Dörner and colleagues (Dörner and Kreuzig, 1983; Dörner, Kreuzig, Reither, and Staudel, 1983) studied individuals who were asked to play the role of city managers for the computer-simulated city of Lohhausen. A variety of problems were presented to these individuals, such as how best to raise revenue to build roads. The simulation involved more than one thousand variables. Performance was quantified in terms of a hierarchy of strategies, ranging from the simplest (trial and error) to the most complex (hypothesis testing with multiple feedback loops). No relation was found between overall cognitive ability and complexity of strategies used. A second problem was created to cross-validate these results. This problem, called the Sahara problem, required participants to determine the number of camels that could be kept alive by a small oasis. Once again, no relation was found between overall cognitive ability and complexity of strategies employed.

The above studies indicate that demonstrated skills do not necessarily correspond between everyday tasks (e.g., price-comparison shopping) and traditional academic tasks (e.g., math achievement tests). In other words, some people are able to solve concrete, ill-defined problems better than well-defined, abstract problems that have little relevance to their personal lives, and vice versa. Few of these researchers would claim, however, that academic skills are totally irrelevant to performance in these various contexts. There is evidence that conventional tests of cognition predict both school performance and job performance (Barrett and Depinet, 1991; Schmidt and Hunter, 1998; Wigdor and Garner, 1982). What these studies do suggest is that there are other aspects of cognition that may be independent of academic cognition and that are important to performance, but that largely have been neglected in the measurement of cognition. We also observe this incongruity between conventional notions of real-world skills in research on age-related changes in cognitive skill.

1.1.2 The fun of growing older: Do age-related patterns in practical cognition resemble those in conventional cognition?

Throughout the century of existence of cognitive psychology, many cognitive variables (mostly those contributing to the *g*-factor—for review, see Berg, in press; Sternberg and Berg, 1992) have been found to be associated with age across the life-span. Most of these associations are rather complex and of curvilinear nature, reflecting rapid growth during the years of formal schooling and slow decline thereafter (Salthouse, 1998). However, the results of research also suggest somewhat different developmental functions for changes in performance on various kinds of cognitive tasks across the adult life span. In particular, data show that older adults commonly report growth in practical skills over the years, even though their academic skills decline (Williams, Denney, and Schadler, 1983).

As for specific cognitive functions, cognition during adulthood is characterized, on one hand, by losses in the speed of mental processes, abstract reasoning, and specific characteristics of memory performance (see Salthouse, 1991, for a review) and, on the other hand, by gains in the metacognitive skill to integrate cognitive, interpersonal, and emotional thinking in a synthetic understanding of the world, self, and others (Labouvie-Vief, 1992, for a review).

The most commonly used theoretical framework adapted for the interpretation of findings on age-related changes in cognitive performance is that of fluid and crystallized cognitive skills (Horn, 1994; Horn and Cattell, 1966). *Fluid* skills are those required to deal with novelty, such as in the immediate testing situation (e.g., discovering the pattern in a figure sequence). *Crystallized* skills are represented by accumulated knowledge (e.g., finding a synonym of a low-frequency word). Utilizing this distinction, many studies have demonstrated that fluid skills are relatively susceptible to age-related decline, whereas crystallized skills are relatively resistant to aging (Dixon and Baltes, 1986; Horn, 1982; Labouvie-Vief, 1982; Schaie, 1977/1978), except near the end of one's life.

In addition, Willis and Schaie (1986) studied the relationships between fluid and crystallized skills and everyday cognition (the latter being defined as the skill to perform core activities of independent life—e.g., cooking, managing finances, or using the telephone and measured by a variant of the ETS Basic Skills Test) in the elderly. The researchers reported substantial correlations between performance on the Basic Skills Test and a measure of fluid (r = .83) and crystallized (r = .78) skills.

The majority of these findings, however, were obtained in the framework of cross-sectional methodologies, that is, by comparing different groups of individuals of various ages. When the same individuals are followed across time in the framework of longitudinal design, the findings show that, with respect to fluid cognition, decline does not generally begin until the sixties and loss of crystallized cognition occurs almost a decade later, in the seventies (Schaie, 1996).

In addition, even when there are age-based group differences in cognitive performance, there is extensive interindividual variability for specific cognitive skills within age groups. For instance, Schaie (1996), although consistently reporting mean cross-sectional differences in overall cognitive performance, pointed out impressive variability within age groups. To quantify this variability, Schaie (1988) investigated the overlap in distributions of cognitive performance among young adults and the elderly. Even in the group of eighty and over the overlap was about 53 percent, scoring well above the mean of their age group. In other words, half or more than half of individuals in the late age groups perform comparably to a group of young adults on measures of both crystallized and fluid cognition.

Moreover, there is also a considerable amount of interindividual variability in the longitudinal patterns of decline, maintenance, and improvement. Specifically, Schaie and Willis (1986) categorized older individuals (the group mean age was 72) into those who decline and those who remained stable in their performances on the Primary Mental Abilities Test (using the space and reasoning subtests) over a period of fourteen years. Forty-seven percent of the sample remained stable on both measures, whereas only 21 percent declined on both measures. Some of these individuals were followed into their eighties, and virtually none of them showed universal descent across all five subtests of the Primary Mental Abilities Test (Schaie, 1989). It is thought that those who show age-related maintenance and improvement in cognitive development differ from those showing decline on a constellation of factors, including educational background, occupational pursuits, health history, life habits, and such personality styles as rigidity and flexibility (Schaie, 1996).

The trend of cognitive development across the lifespan, however, appears to be yet somewhat different for practical skills. Williams et al. (1983) interviewed men and women over the age of 65. The questions posed to these adults had to do with their perception of age-related changes in their skill to think, reason, and solve problems. Surprisingly enough, the responses obtained from these adults were largely contradictory to the view that late development of cognition consists of decline (see Berg, in press, for review). In the Williams et al. study (1983), 76% of the elderly adults believed that their skill to think, reason, and solve problems had actually increased over the years, with 20% reporting no change and only 4% reporting that their skills had declined with age. The researchers confronted the participants with the overwhelming evidence of decline in conventional test performance upon completion of formal schooling, but the explanation of the elderly people was that they were talking about solving kinds of problems different from those found on psychometric tests. The problems they had in mind when answering the interviewer's questions were those of an everyday or financial nature. Of course, these responses might be simply discounted as self-deceiving and self-reassuring, but a number of formal psychological studies within the last decade have provided significant support for the claim made by the elderly in the Williams et al. (1983) study.

In particular, the idea that practical and academic skills might have different developmental trajectories was supported in a number of studies (see Berg and Klaczynski, 1996, for a review). Denney and Palmer (1981) were one of the first research teams to demonstrate this discrepancy. They compared the performance of adults (aged 20 through 79) on traditional analytical reasoning problems (e.g., a "twenty questions" task) and a problem-solving task involving real-life situations (e.g., "If you were traveling by car and got stranded out on an interstate highway during a blizzard, what would you do?"). One of the many interesting results obtained in this study was a difference in the shape of the developmental function for performance on the two types of problems. Performance on the *traditional* problem-solving task or cognitive measure declined almost linearly from age 20, onward. Performance on the *practical* problem-solving task increased to a peak in the 40- and 50-year-old groups, declining thereafter. Expanding on this line of research, Smith and colleagues (Smith, Staudinger, and Baltes, 1994) compared responses to life-planning dilemmas in a group of younger (mean age 32) and older (mean age 70) adults. Unlike the results of studies of aging and academic skills, which demonstrated the superior performance of younger adults over the elderly, in this study, young and older adults did not differ. In addition, each age-cohort group received the highest ratings when responding to a dilemma matched to their own life phase.

Similar results were obtained in a study by Cornelius and Caspi (1987). They studied adults between the ages of 20 and 78. These researchers examined relationships between performance on tasks measuring fluid cognition (letter series), crystallized cognition (verbal meanings), and everyday problem solving (e.g., dealing with a landlord who won't make repairs, filling out a complicated form, responding to criticism from a parent or child). Performance on the measure of fluid skill increased from age 20 to 30, remained stable from age 30 to 50, and then declined. Performance on the everyday problem-solving task and the measures of crystallized skill increased through age 70.

Likewise, the neofunctionalist position, advanced by Baltes and his associates (Baltes, 1987; Baltes et al., 1984; Baltes, Smith, and Staudinger, 1992; Dittmann-Kohli and Baltes, 1990) acknowledges that, although some aspects of cognitive functioning estimated via traditional tests may decline with age, stability and growth also exist, if to a lesser extent. The approach of Baltes and his colleagues also operates within the constructs of fluid and crystallized cognition, although a different emphasis is placed on the relative roles and meanings of these two kinds of cognition. Here, both aspects of cognition are considered as coequals in defining the developmental course of cognition. In general, Baltes argues that crystallized cognition has been too narrowly defined, and that its importance increases as one moves into adulthood and old age. In this sense, it may be inappropriate to associate a decrease in fluid cognition with an average decline in cognitive competence. Baltes and his associates see adult cognitive competence in terms of a dual-process model. The first process, called the *mechanics* of cognition, is concerned with developmental change in basic information processing that is genetically driven and assumed to be knowledge-free. With aging, there is a biologically-based reduction in reserve capacity (Baltes, 1987; Baltes et al., 1992). The second process, *pragmatic* cognition, relates the basic cognitive skills and resources of the first process to everyday cognitive performance and adaptation. Measures of pragmatic cognition within select domains are viewed as tapping skills more characteristic of adult cognitive life than are traditional psychometric measures of cognitive skills. Similar to empirical findings on the distinction between fluid and crystallized cognition, Baltes, Sowarka, and Kliegl (1989) showed that the mechanics of cognition tend to decline with age almost linearly, whereas the pragmatics of cognition tend to maintain relative stability throughout adulthood. For example, whereas linear declines were found in the speed of comparing information in short-term memory (i.e., aspects of cognitive mechanics), no age differences were registered for measures of reasoning about life planning (i.e., aspects of cognitive pragmatics). Cognitive skills are assumed to operate on content domains involving factual and procedural knowledge; they are regulated by higher-level, trans-situational, procedural skills and by higher-order reflective thinking (metacognition), all of which define the "action space" in which problem solving occurs within a given individual. According to this approach, successful aging entails limiting one's tasks and avoiding excessive demands. Baltes and Baltes (1990) use the concept of selection to refer to a self-imposed restriction in one's life to fewer domains of functioning as a means to adapt to age-related losses. It is assumed that by concentrating upon high-priority domains and devising new operational strategies, individuals can optimize their general reserves (Baltes, 1993). By relating adult cognition to successful cognitive performance in one's environment, this position acknowledges that not all tasks are equally relevant for measuring cognition at different ages (Baltes et al., 1984; Baltes et al., 1992).

Specific manifestations of pragmatic cognition are said to differ from person to person as people proceed through selection, optimization, or compensation (Dittmann-Kohli and Baltes, 1990). Selection refers simply to diminishing the scope of one's activities to things that one is still able to accomplish well, despite a diminution in reserve capacity. Thus, research shows that elderly people tend to leave jobs that require quick sensorimotor responses (Barrett, Mihal, Panek, Sterns, and Alexander, 1977). Optimization refers to the fact that older people can maintain high levels of performance in some domains by practice, greater effort, and the development of new bodies of knowledge. Compensation comes into play when one requires a level of capacity beyond remaining performance potential. For example, Salthouse (1984) was able to show that older typists, although slower on several simple speeded reaction-time tasks, were able to compensate for this deficit and maintain their speed by reading further ahead in the text and planning ahead. According to Salthouse and Somberg (1982), age-related decrements at the "molecular" level (e.g., in speed of execution of the elementary components of typing skill) produce no observable effects at the "molar" level (i.e., the speed and accuracy with which work is completed).

Charness (1981) showed similar effects with older chess players, who exhibited poorer recall in general, but were better able to plan ahead than younger, less experienced players. In related studies, older adults have been found to compensate for declines in memory by relying more on external memory aids than do younger adults (Loewen, Shaw, and Craik, 1990). Older adults must often transfer the emphasis of a particular task to skills that have not declined in order to compensate for those that have (see Bäckman and Dixon, 1992, for a review of these issues). In other words, when a task depends heavily on knowledge, and speed of processing is not a significant constraint, peak performance may not be constrained in early-to-middle adulthood (Charness and Bieman-Copland, 1994). As an example, consider chess competitions by correspondence. In these "chess-by-mail" competitions, players are permitted three days to deliberate each move. The mean age of the first-time winners of one postal world championship is 46 years old. In contrast, the peak age for tournament chess, where deliberation averages three minutes per move, is about 30 years old (Charness and Bosman, 1995).

A series of studies on the relationship between aging and cognitive efficiency in skilled performers attested to the compensatory and stabilizing role of practical cognition (Baltes and Smith, 1990; Charness and Bosman, 1990; Colonia-Willner, 1998; Hartley, 1989; Willis, 1989). Sternberg and colleagues' studies of tacit knowledge in the domains of business management, sales, and academic psychology showed increases in tacit knowledge with age and experience across groups of undergraduates, graduate students, and professionals (Sternberg, Wagner, Okagaki, 1993; Wagner, 1987; Wagner, Rashotte, and Sternberg, 1994; Wagner and Sternberg, 1985). Colonia-Willner (1998) found evidence that older managers who performed at the highest levels on average had high levels of tacit knowledge—even though on average they had relatively low scores on psychometric reasoning measures. In addition, Colonia-Willner pointed out an interesting detail: even though tacit knowledge of managerial skills was shown to be related to some indicators of job success for the total sample of bank managers, the relative weight of this knowledge was higher for the highest success group (that group rewarded most highly). It might be that job-related tacit knowledge is especially important for detecting super-achievers among a fairly restricted, high-achieving, conventional population of managers engaged in heterogeneous activities.

Moreover, a series of training studies, conducted in Germany (Baltes et al., 1984; Baltes et al., 1992) and the U.S. (Schaie, 1986; Schaie and Willis, 1986; Willis and Schaie, 1994), have shown that older individuals still have a great deal of potential plasticity, or reserve capacity for development. The results demonstrated that intervention can lead to significant gains in skills such as problem-solving tasks (Denney, 1979), perceptual speed (Hoyer, Labouvie, and Baltes, 1973), and fluid cognition (Baltes and Lindenberger, 1988; Willis, 1987). Intervention research generally targeted those skills which have been shown to decline the most (i.e., fluid cognition and processes representative of the mechanisms of cognition).

In general, results from intervention studies convincingly demonstrated the remarkable plasticity of human cognition in the elderly (see Willis, 1987 for a review). In the German studies, better performance was demonstrated for (1) target training (Baltes and Willis, 1982; Willis, Blieszner, and Baltes, 1981), (2) independent self-practice (Baltes et al., 1989; Hayslip, 1989a, 1989b), and (3) removed time constraints (Hofland, Willis, and Baltes, 1981). Willis and Schaie, 1986; Schaie and Willis, 1986; Willis and Schaie, 1994) obtained similar findings within a longitudinal design.

These results were replicated in a second follow-up study conducted in 1991 with both new participants and participants from the original training study. Specifically, results from the Seattle Training Study, a component of the Seattle Longitudinal Study (Schaie, 1996) indicated that the performance of the elderly can be successfully impacted in such a way that older adults' performance is boosted back to the level at which they performed more than a decade before. The Seattle researchers set up five one-hour sessions aimed at training the elderly adults' spatial and reasoning skills. The training had differential impact on certain subgroups of the elderly population. For those who had shown decline on either of the Primary Mental Skill Test subtests over the preceding fourteen-year period, training was effective in returning their performance nearly to the original level. For those who had remained stable over the preceding fourteen-year period, training raised their performance beyond the level they performed at fourteen years prior to the training. In addition, the training has been found to be effective, not only in the short run, but over seven years (Neely and Backman, 1993; Willis and Nesselroade, 1990).

One of the outcomes of these studies is the realization that longer and more structured training seems to be necessary for remediation in the very old (Schaie, 1994; Willis, 1989). The importance of these studies is that they suggest that cognitive decline in many individuals may be due to disuse of certain cognitive skills, and that remediation is possible for a significant number of participants, especially for the young-old (Schaie, 1994; Willis, 1990; Willis and Schaie, 1994).

The developmental trajectory of everyday cognition has been examined by a number of researchers (see Berg, in press; Berg and Klaczynski, 1996, for review). The summary of the field today is that the pattern of age differences in everyday cognition differs dramatically depending on how problems to be solved are defined and what criteria are used for optimal problem solving. For example, Berg, Klaczynski, Calderone, and Strough (1994), studying participants' own ratings of how effective they were in solving their own everyday problems, did not find any age differences. Denny and her colleagues (Denney and Palmer, 1981; Denney and Pearce, 1989) utilized the number of "safe and effective solutions" as the criterion for optimal problem solving and found that the highest number of such solutions was generated by middle-aged adults, with both younger and older adults offering fewer solutions. Cornelius and Caspi (1987), using the closeness between participants' ratings of strategy effectiveness and a "prototype" of the optimal everyday problem solver as the criteria, found an increase in everyday problem-solving skill with adult age.

A number of studies have examined everyday problem solving with a neo-Piagetian approach to cognitive development in adulthood (Labouvie-Vief, 1992). According to this paradigm, in middle and late adulthood, the formal operational reasoning of late adolescents and young adults, with its focus on logic, is replaced by more sophisticated mental structures distinguished by relativistic reasoning based on synthesizing the irrational, emotive, and personal. Specifically, Blanchard-Fields (1986, 1994; Blanchard-Fields, and Norris, 1994) stated that, when dealing with social dilemmas, older adults are superior to younger adults in their integrative attributional reasoning (i.e., reasoning based on the integration of dispositional and situational components).

To conclude, there is reason to believe that the developmental trajectories of skills utilized to solve strictly academic problems do not coincide with the trajectories of skills used to solve problems of a practical nature.

1.2 What develops in practical cognition?

The evidence supporting the supposition that practical cognition has a different developmental trajectory than academic cognition supports the etiological independence (not necessarily complete) of practical and academic skills but is only one of many research advances revealing the developmental mechanisms of practical cognition. Developmental research on practical skills is still in its early stages. However, data available at this point shed some light on what Sinnott (1989) called the *chaotically complex* reality of practical problem solving; evidence supports the existence of different developmental trajectories (maintenance, improvement, and decline) across the life span without a pronounced preference for any single one.

There is no formal theory of the stages of the development of practical cognition (Berg, 1994). Some results, however, suggest that the difference in performance on practical and analytical tasks is observed rather early. Freeman, Lewis, and Doherty (1991) have shown that the performance of preschoolers on the false-belief tasks (e.g., tasks involving the formation of false beliefs and expecting children to determine and overcome their false nature) is better if they are asked to act out answers rather than to give them verbally. The researchers suggest that the reason for this discrepancy is that early implementation of a theory of intentionality is “only” practical. In other words, preschool children are able to distinguish between true and false expectations and true and false causes, but do it by carrying out practical actions (e.g., acting with the right object) rather than by explaining why those particular objects should be chosen. These and other findings contribute to the hypothesis that reflective awareness and verbalization emerge gradually from the implicit practical cognition organizations which are their necessary precursors (e.g., Bickhard, 1978; Karmiloff-Smith, 1988).

Developmental research on practical cognition is moving in a number of directions, each of which might help us to detect the internal mechanisms of its development. Most of the work is centered on specific characteristics of practical tasks. The assumption here is that if we understand the differences in the ways these tasks are formulated and solved at different stages of development, we will be closer to understanding the developmental dynamics of practical cognition. Drawing on the distinction made earlier between academic and practical tasks suggests five main directions of research: (1) studies of developmentally variable contexts of practical problem solving; (2) studies of developmental changes in the content of practical problems encountered at different stages of development; (3) studies of the developmental diversity of the goals of practical problem solving; (4) studies of differential strategies utilized in practical problem solving at different periods of development; and (5) studies on developmental variation in problem interpretation and definition.

1.2.1 Context of practical problem solving

There is virtually unanimous agreement on the centrality of context for understanding practical problem solving. This view, which holds that practical problem solving cannot be separated from the context in which it unfolds, is referred to as the contextual perspective (e.g., Dixon, 1994; Wertsch and Kanner, 1994). In general, the metaphor used to describe the contextual approach is that of trying to follow forever changing events (i.e., the life course is represented as being a series of changing events, activities, and contexts). When applied to studies of practical problem solving, this perspective assumes that (1) the demands posed by these contexts vary across development; (2) strategies accomplishing adaptation differ across contexts; (3) these strategies also differ across individuals; and, finally, (4) the effectiveness of everyday problem solving is determined by the interaction of individual and context (Berg and Calderone, 1994). Several studies have found that the context in which the problem occurs (e.g., family, work, or school) impacts everyday problem solving in all its components (content, goal, and strategy).

Consider the following examples. Ceci and Bronfenbrenner (1985; Ceci, 1990), employing a dual context paradigm, have conducted a series of studies concerning the impact of physical and social contexts on cognition. The dual context paradigm proposes that children be made to perform the same task in two or more contexts. The assumption here is that some settings elicit more effective forms of cognition than do others by stimulating or activating different strategies. The Ceci-Bronfenbrenner view is that a task perceived in a modified form might recruit a set of strategies acquired previously but not elicited by the original, unmodified task. (For example, a video-game task, which is a modification of a simple task requiring a participant to follow the movements of dots, might recruit strategies that the dot task alone would not.) Cohen (1996) studied the mathematically-oriented activity of 3- and 4-year olds and found that, when mathematical operations were embedded in the broader context of a "play-store" setting, children were able to solve problems that exceeded an age-appropriate level of difficulty. In addition, the children satisfied the demands of the task in using a variety of solution strategies.

One of the most interesting developments in studies on context and practical problem solving concerns the effect of compensation: the phenomenon in which gains in (mostly) practical cognition balance out age-related decrements in others. Researchers argue that compensation—considered in terms of the dynamic relationship between the individual's changing cognitive skills and expectations of performance, on the one hand, and shifting contextual demands, on the other hand—should be viewed as central to cognitive aging (e.g., Dixon, 1994). One example of practical cognition compensating for declines in *g*-based cognitive performance is older adults' effective use of external aids. One common source of external cognitive aid is other people. For example, Dixon and his colleagues (Dixon, 1994) explored the extent to which older and younger adults use same-age collaborators in solving memory problems and found that older adults use previously unknown collaborators to boost their performance levels to a much greater extent than do younger adults.

Two other important characteristics of the context in which practical problem solving occurs, which might explain some aspects of the observed development variability in practical cognition, are the complexity and familiarity of the context.

As for the complexity of the environment in which practical cognition unfolds, one variable that has been pointed out as extremely important for shaping the development of practical skills in adulthood is that of the immediate conditions and demands of work (see Schooler, in press, for a review). For example, Kohn and Schooler (1983), in a group of men between the ages of 24 to 64, longitudinally studied the link between the extent to which one's work-related activities involve independent thought and judgment and workers' flexibility in dealing with complex cognitive demands. They found that the more the substantive complexity of one's job, the greater the incremental gains in cognitive performance over a ten-year period. Even more astounding, a similar relationship between job complexity and cognitive performance was revealed for women doing complex housework (Schooler, 1984). Moreover, K.A. Miller and Kohn (1983) found that individuals with higher flexibility in dealing with complex cognitive activities tended to engage in more stimulating and demanding cognitive activities (e.g., reading books versus watching television). The major criticism of this nonexperimental evidence of the cognitive effects of doing complex work (whether in the work place or the household) is that these designs are unable to rule out the possibility that individuals who maintain their cognitive functioning are more capable of pursuing and staying in challenging work environments. Yet, even though the causal path is difficult to infer among individuals, the evidence that among individuals more cognitively complex work leads to enriched cognitive functioning deserves attention and more thorough investigation.

Regarding familiarity or experience with the domain in which practical problem solving is carried out, studies have demonstrated that cognitive performance is greater for both young and older adults when individuals are given either familiar materials (Smith and Baltes, 1990) or a chance to practice prior to assessment (Berg, Hertzog, and Hunt, 1982). Yet, results are ambiguous as to whether differential familiarity is a factor that can help to explain age differences in practical problem solving (Denney and Pearce, 1989).

Researchers reported, for example, that older adults perceived traditional cognition tests as less familiar than did young adults (Cornelius, 1984). Therefore, when younger and older adults are compared on conventional cognition tests, older adults might look worse because these tests are less familiar to them and they may have forgotten how to evoke specific strategies relevant to situations of cognitive assessment.

To explore the importance of the familiarity factor, several studies have been carried out in which younger and older adults were asked to solve problems that were constructed to be more familiar or more normative for one age group or the other. For example, Denney and colleagues (Denney, Pearce, and Palmer, 1982) showed that, in adults, the more normative for their age group everyday problems are, the better their performance is. Similarly, Smith and Baltes (1990) found that adults perform best when the problems are more normative for their age group. As Berg (in press) pointed out, memory research utilizing the usage of tasks with familiar materials (e.g., remembering words that were in frequent use during their adulthood years versus contemporary equivalents) is consistent in showing that older adults tend to perform better with materials more familiar to them (Barret and Watkins, 1986; Worden and Sherman-Brown, 1983).

1.2.2 Content of practical problem solving

The main hypothesis underlying this line of research is that the content of practical problem solving differs at different stages of development. The literature published to verify this hypothesis contains heterogeneous evidence; some is supportive (e.g., Aldwin, Sutton, Chiara, and Spiro, 1996) and some is not supportive (e.g., Folkman, Lazarus, Pimley, and Novacek, 1987) of the assertion that individuals of different ages experience different everyday problems.

Berg and colleagues (Berg and Calderone, 1994; Sansone and Berg, 1993) asked preschoolers, teenagers, college students, and older adults to describe a recent problem (hassle, conflict, challenge, and so on) that they had experienced and to describe the problem in as much detail as possible. The intent was to investigate whether the types of domains of problems remain constant across development or whether different types of problems would appear for different age groups. The researchers found significant variation in the content of everyday problems across development. The everyday problem-solving content for 5-6-year-olds consisted predominantly of problems dealing with family (e.g., disagreements with family members) and assigned responsibilities (e.g., home chores). For 11 to 12-year-olds, everyday life problems centered on school and after-school activities and environments. No single content area dominated the everyday life of college students, and their salient problems had to do with free time, work, friends, family, and romantic relationships. Finally, the everyday problem solving of the older adults centered on the family context and health.

Barker (1978) suggested that the content of practical problem solving is determined by the ecological characteristics of a given developmental period. They carried out detailed observations of settings inhabited and experienced by elementary school children on a daily basis and found that children most frequently occupy settings embedded in schooling and family life. This piece of work is unique in terms of its thoroughness and attention to details; however, based on sporadic evidence accumulated in research on developmental life tasks, the general assumption in the field is that the content of the practical problem solving of adults differs in a variety of ways across the life span. In other words, it might be impossible to carry out Baker et al.-like studies in all ecological settings encountered in adulthood, but it might be possible to target the few that appear to be crucial at specific developmental periods. Specifically, it has been shown that (1) college students' tasks are primarily aimed at succeeding academically, forming social networks, developing an identity, and separating from family (Cantor, Norem, Neidenthal, Langston, and Brower, 1987); (2) adults focus on a variety of tasks, ranging from starting a family and a career in young adulthood, through the pragmatic tasks of middle adulthood, to adapting to impairments of health and adjusting to retirement during old and advanced old age (Baltes et al., 1984; Havinghurst, 1972; Neugarten Moore, and Lowe, 1968).

1.2.3 Goals of practical problem solving

The goal-directedness (e.g., Goodnow, 1986; Scribner, 1986; Wertsch, 1985) of practical problem solving is one of the most often cited characteristics of practical cognition in application. Therefore, the second line of research concerns the developmental trajectories of goals of practical problem solving.

Strough, Berg, and Sansone (1996) showed that there is developmental variation in the types of goals underlying everyday problem solving. The profile of this developmental variation reflects developmental life tasks (Cantor, 1990). Specifically, preadolescents reported more goals for task improvement, and a large portion of their problems involved the school context. Interpersonal goals appeared to be more salient to middle-aged adults than to preadolescents. Preadolescents, however, reported more other-focused assistance-recruiting goals than did adults. Older and middle-aged adults reported more physical goals than did younger individuals, and the adult group as a whole reported more affective goals than did preadolescents.

Klaczynski, Laipple, and Jurden (1992) studied practical cognition among adolescents in college-preparatory or vocational-training tracks. Depending on the chosen developmental life-track, adolescents in the two groups differed in their interpretation of practical problem situations. In particular, vocational students were concerned primarily with goals involving the acquisition of adult status, such as marriage, steady employment, and independence. College-preparatory students, on the other hand, reported more achievement-oriented goals, such as doing well in school, gaining admission to quality colleges, and scoring well on entrance exams.

Belief in the plasticity and fluidity of human developmental goals throughout the life span is also reflected by the notion that there is no single outcome or endpoint to cognitive development in general, or to the development of practical cognition in particular (e.g., Rogoff, 1982). The implication of this line of reasoning is that the individual and his or her context form a complex systemic unit; changes in the unit shape the content, dynamics, and adaptability of the individual's cognitive functioning in specific contexts. Thus, there is no "ideal" trajectory of cognitive development, and there is no optimal instrument assessing cognitive functioning equally well at all periods of the life span.

1.2.4 Practical problem-solving strategies

One of the main research trajectories in the field of practical cognition focuses on strategies utilized in problem solving. Among the central characteristics of strategies discussed in the research literature of the past 20 years (Belmont and Butterfield, 1969; Berg, 1989; Brown, 1975; Flavell, 1970; Naus and Ornstein, 1983; Pressley, Forest-Pressley, Faust and Miller, 1985) are selectivity, goal-directedness, and intentionality. Many developmental researchers have been especially interested in strategy selection as both an individual and a developmental indicator of everyday problem-solving performance (e.g., Frederiksen, 1986; Frederiksen, Jensen, and Beaton, 1972; Lazarus and Folkman, 1984).

Most of the early developmental work on everyday problem solving has been carried out under the assumption that individuals' chosen strategies can be compared irrespective of the developmental variation in the goals motivating these strategies (Band and Weisz, 1988; Berg, 1989; Cornelius and Caspi, 1987; Folkman et al., 1987). The major theoretical hypothesis dominating the field is that greater experience with everyday problems leads to better problem solving (Baltes et al., 1984, Denney, 1982). This claim assumes that a particular type of strategy—e.g., primary control reflected in independent coping and problem-focused action—is a more effective way of dealing with various problems than is some other strategy—e.g., secondary control reflected in reliance on others and emotion-focused action (Denney, 1989; Folkman et al., 1987). For example, self-action was the strategy most frequently mentioned across all ages in a study of reported everyday problems (Berg, Strough, Calderone, Sansone, and Weir, 1998). Problem-focused action was most frequently mentioned for hypothetical problems (Blanchard-Fields, Jahnke, and Camp, 1995). Developmental differences have been encountered, suggesting that secondary control strategies, emotion-focused strategies, and dependence on others increases across early childhood (Band and Weisz, 1988), with further elevation in later adulthood (Brandtstaedter and Greve, 1994; Denney and Palmer, 1981; Folkman et al., 1987; Heckhausen and Schultz, 1995). For instance, researchers (Band and Weisz, 1988) found that older children were more likely to use secondary control strategies, such as efforts to modify the subjective psychological state of the self to better suit the present conditions of the problem, whereas younger children were more likely to use primary control strategies, such as efforts to influence the problem so that it meets the problem solver's expectations.

The empirical literature, however, does not uniformly support the claim that "more experience equals better problem solving" (Baltes, 1997; Berg, 1989; Cornelius and Caspi, 1987). Recent research suggests that strategies are differentially effective depending on the context of the everyday problem (Berg, 1989; Ceci and Bronfenbrenner, 1985; Cornelius and Caspi, 1987; Scribner, 1986). Thus, Cornelius and Caspi (1987) showed that different types of strategies (problem-focused action, cognitive problem analysis, passive-dependent behavior, and avoidant thinking and denial) were viewed as differentially effective in different contexts.

Findings regarding the localization of age differences are also somewhat contradictory. The often-cited trend in the literature is that older adults tend to use more secondary control (e.g., Heckhausen and Schulz, 1995) and less problem-focused action or primary control (Folkman et al., 1987) when compared to younger adults. Blanchard-Fields et al. (1995) found minimal age differences in problem-focused action. Furthermore, Berg et al. (1998) reported age differences for older adults only, with older people using relatively less cognitive regulation and more self-action than either college students or middle-aged adults. The situation has become even less transparent, with Aldwin et al. (1996) showing that, for the most part, age differences existed among adults only when individuals' strategies were assessed through a checklist; these distinctions were greatly reduced when individuals' strategies were elicited through open-ended interviews.

One of the possible explanations for the heterogeneity of these findings is that what develops over time is sensitivity to specific contexts. In other words, the repertoire of dealing with everyday problems is rather broad, and different modules of problem solving are used in different situations; in many ways, consistency across situations may be maladaptive (Mischel, 1984). Some researchers argue that successful everyday problem solving will involve carefully fitting strategies to the specific demands of a problem and modifying these strategies in response to changes in the problem (Berg and Sternberg, 1985; Rogoff, Gauvain, and Gardner, 1987; Scribner, 1986). And sensitivity to the contextual features of a problem is characteristic of a developmental factor (Mischel, 1984; Rogoff et al., 1987). Others, on the contrary, suggest that these strategies become less context-dependent with age (e.g., Kreitler and Kreitler, 1987).

Yet another, although not contradictory possibility, is that the lesson derived from experience with everyday problems is how to avoid getting into everyday problems (Berg, 1989). Thus, it is plausible that no simple relation between kind of experience and everyday problem-solving skill is likely to exist. Moreover, researchers have presented evidence demonstrating that so-called effective-across-all-contexts (e.g., primary) strategies fail in situations in which so-called ineffective strategies (e.g., relinquishing) work (Berg, Calderone, and Gunderson, 1990, as cited in Berg and Calderone, 1994). Certain kinds of experience may be differentially related to success at solving particular kinds of everyday problems, and development might better be construed as individuals becoming increasingly capable of modifying their strategies or avoiding potentially problematic situations (Berg, 1989; Rogoff et al., 1986).

Another line of research focuses on studying individual differences that appear to lead to more optimal problem-solving performance (e.g., Ceci and Liker, 1986; Denney, 1989; Willis and Schaie, 1986). Many factors (e.g., conventional cognitive skills, personality traits, social skills, achievement motivation) have been shown to impact the utilization of strategies in everyday problem solving (e.g., Ceci and Liker, 1986; Charness, 1981; Kuhn, Pennington, and Leadbeater, 1983), but no specific constellations of these factors were found to be better predictors of effective problem solving.

1.2.5 Problem interpretation (Definition)

In an attempt to systematize the literature on the development of everyday problem solving, Berg and colleagues have introduced the concept of "problem interpretation" (Berg and Calderone, 1994; Sansone and Berg, 1993) or "problem definition" (Berg et al., 1998). The problem interpretation arises at the edge of the context and the individual and, in essence, is the transaction of the individual with his or her context. The problem interpretation derives from features of both the individual and the context, but it might selectively engage all or only some features. Berg and her colleagues argue that such individual and contextual features may have different weights and may be differentially combined at different stages of development; thus, the search for developmental variation in everyday problem solving should focus on the development of problem interpretation (Berg and Calderone, 1994).

As it is interactive in nature, problem definition reflects those aspects of the self and context that are activated with respect to a specific problem unfolding at a specific moment in time. Problem definition is a complex, psychological, subjective reality, which, according to Berg et al. (1998), reflects the individual's goals and expectations (Bandura, 1986), determines the strategies to be used to meet these expectations and accomplish subjective goals (Vallacher and Wegner, 1987), affects the outcome attribution and meaning interpretation (Dodge, Pettit, McClaskey, and Brown, 1986), and induces the affective representation of the problem (Fleeson and Cantor, 1995).

A number of studies provide supportive evidence for the transactional approach to everyday problem solving. Sinnott (1989) showed that older adults' interpretation of Piagetian logical-combination problems, especially those experienced in real life (e.g., assigning relatives to sleeping locations), vary to a greater degree than do the interpretations of younger adults. Specifically, older adults tend to be more sensitive to social and interpersonal facets of the problem when compared with younger adults, who concentrate on the problem's logical aspects. Similarly, Laipple (1992) showed that older adults were less likely to interpret the situation of solving logical problems with the meaning intended by the experimenter; older adults tended to leave the logical confines of the problem and inject into the experimental situation more personal experience than did the younger adults. Chi and Ceci (1987) suggested that many types of problem solving appear to be directly influenced by the mental context the child brings to the task.

In their own work, Berg and colleagues (Berg and Calderone, 1994) registered a number of developmental characteristics of problem definition. First, they showed that, with age, there was a decrease in the frequency of task-oriented interpretations of problems and an increase in interpersonal, self, and mixed (e.g., task and self) interpretations. In their interpretation, researchers suggest that these findings correspond to the literature on the development of the self system, according to which changes of the self system involve movement away from a concrete and specific system to one that incorporates more abstract and interrelated psychological constructs (Harter, 1983). Second, Berg et al. (1998) studied the link between the problem definition and the selection of strategies for problem solving. In general, problem definition appears to be a more precise predictor of strategy use than does problem context. Specifically, individuals who defined a problem in terms of interpersonal concerns alone were more likely to report using strategies involving regulating or including others. On the contrary, individuals who defined a problem solely in terms of competence concerns were more likely to utilize strategies including independent action and less likely to engage others. Finally, the links between problem definition and strategy selection were not found to vary as a function of age.

Problem definition is very important to practical cognition. For example, a key difference between the results of Berg et al. (1998) and those of previous research is the importance that individuals placed on the social aspects of practical problem solving. Berg and colleagues found that the majority of individual problem definitions in any age group (pre-adolescents, college students, and adults) involved interpersonal concerns. These problem definitions, in turn, determined the selection of strategies that involved regulating or including others. Note that this interpretation differs significantly from the argument utilized in previous research. Earlier work typically assumed that reliance on others reflected ineffective problem solving because individuals exhibited dependence on others (e.g., Cornelius and Caspi, 1987; Denney and Palmer, 1981; Folkman et al., 1987). However, the reinterpretation of the role of social-dependent strategies suggests that using others to deal with everyday problems is a strategy rather well suited to particular problems (Baltes, 1997; Meacham and Emont, 1989).

2. Approaches to studying practical cognition

During the past two decades, there has been a growing interest (and in part a renewed interest) in nonacademic forms of cognition. Several distinct, but arguably overlapping, constructs have been proposed to capture this nonacademic form of cognition. One of these constructs is Sternberg's (1985a, 1997a) concept of practical cognition. Alternative related conceptualizations of nonacademic or practical cognition include *social cognition* (e.g., Cantor and Kihlstrom, 1987, Ford and Maher, 1998; Kihlstrom and Cantor, in press), *emotional cognition* (e.g., Goleman, 1995; Salovey and Mayer, 1990; Mayer, Salovey, and Caruso, in press), and *intrapersonal* and *interpersonal cognitions* (Gardner, 1983, 1993). Jones and Day (1997) noted the similarities among the various conceptualizations of nonacademic cognition. They suggested that practical, social, and emotional cognition share a focus on declarative and procedural knowledge, flexible knowledge-retrieval capabilities, and problem solving involving more than one correct interpretation or solution. We discuss the different conceptualizations of practical cognition and the methods researchers have used to study them.

2.1 Social cognition

Interest in the construct of social cognition has fluctuated since the concept was first introduced by Thorndike (1920). Thorndike defined social cognition as comprising the skills to understand others and to act or behave wisely in relation to others. He also distinguished social from abstract and mechanical forms of cognition. Several other definitions and expansions on Thorndike's definition followed. These expanded definitions included the skill to get along with others (Moss and Hunt, 1927), the skill to deal with people (T. Hunt, 1928), knowledge about people (Strang, 1930), ease with other people, insights into the states and traits of others (Vernon, 1933), and the skill to judge correctly the feelings, moods, and motivations of others (Wedeck, 1947). Wechsler's (1958) definition seemed to capture these various conceptualizations in the single definition of social cognition as one's facility in dealing with human beings.

Some researchers sought to understand the meaning of social cognition by studying people's implicit concepts or theories (e.g., Bruner, Shapiro, and Tagiuri, 1958; Cantor, 1978). In a study by Sternberg et al. (1981), discussed previously, experts and laypersons were asked to rate how characteristic various behaviors were of intelligent, academically intelligent, and everyday intelligent people. A factor of "social competence" emerged from the factor analyses of the ratings in each aspect of cognition.

More recently, Kosmitzki and John (1993) attempted to clarify some of the inconsistency in the literature regarding definitions of social cognition. They identified seven components that seemed to be most central to people's implicit conceptions of social cognition. The seven components included both cognitive elements (perspective taking, understanding people, knowing social rules, and openness to others) and behavioral elements (good at dealing with people, social adaptability, and interpersonal warmth). These implicit conceptions overlap, to some extent, with scientists' explicit theories, but suggest some additional aspects previously not included, such as interpersonal warmth and openness to others. Although these last two aspects have yet to be tested empirically, most studies have focused on some variation of the five remaining components (perspective taking, understanding people, knowing social rules, skill to deal with people, and social adaptability).

Throughout its history, the study of social cognition has periodically fallen out of favor with researchers. This lack of interest can be attributed to failed attempts to distinguish measures of social from measures of abstract cognition. The difficulty in distinguishing social from academic or abstract cognition can be explained by efforts that focus primarily on cognitive aspects of social cognition and methods that rely heavily on verbal assessment. Researchers as early as Thorndike (1920) acknowledged the multidimensional nature of social cognition. Until recently, however, the approaches to studying social cognition have emphasized cognitive aspects, such as social perception (e.g., Chapin, 1942) and moral reasoning (e.g., Keating, 1978). In order to assess these cognitive dimensions, researchers relied, to a large extent, on verbal measures. Measures of behavioral aspects of social cognition also have relied somewhat on verbal forms of assessment (e.g., self-report). As becomes clear from a brief review of the literature, research efforts that consider behavioral and nonverbal measures of social cognition have had greater success in establishing discriminant validity from measures of abstract cognition than have the more cognitive, verbal measures of social cognition.

2.1.1 Cognitive-verbal measures of social cognition

Many approaches to understanding social cognition follow the tradition of cognition testing by developing instruments to assess individual differences in social cognition. One of the first and better known tests of social cognition was the George Washington Social Cognition Test (GWSIT; Moss, Hunt, Omwake, and Woodward, 1949). This test consists of a number of subtests that assess judgment in social situations, recognition of the mental states behind messages, memory for names and faces, observation of human behavior, and sense of humor. Early research with the GWSIT suggested that it could not be distinguished easily from abstract cognition (e.g., Thorndike and Stein, 1937).

A set of social-cognition tests emerged within the context of Guilford's (1967) Structure of Intellect Model of Cognition. Within Guilford's framework, social cognition is viewed as comprising those skills within the domain of behavioral operations. O'Sullivan, Guilford, and deMille (1965) developed tests to measure behavioral cognition, which they defined as the skill to judge people. More specifically, the tests measured the skill to decode social cues, including facial expressions, vocal inflections, posture, and gestures. In a study with 306 high-school students, O'Sullivan et al. (1965) found evidence that their factors of social cognition were distinct from measures of abstract cognitive skill. Later research, however, found contradictory results (e.g., Riggio, Messamer, and Throckmorton, 1991).

Riggio et al. (1991) administered several measures of social cognition and several measures of academic cognition to undergraduate students. Academic cognition was measured using the Shipley-Hartford Institute of Living Scale (Shipley, 1940), which measures verbal and abstract reasoning, and the vocabulary subscale of the WAIS-R (Wechsler, 1981). Measures of social cognition included four tests of the Factor Tests of Social Cognition (O'Sullivan and Guilford, 1976); Riggio's (1986, 1989) Social Skills Inventory (SSI), which assess six social communication skills (emotional expressivity, emotional sensitivity, emotional control, social expressivity, social sensitivity, and social control); and a social etiquette/tacit knowledge test that measured knowledge of appropriate behaviors in social situations. Riggio et al. found comparable intercorrelations within measures of both academic and social cognition as they did between measures of academic and social cognition. An exploratory factor analysis suggested two factors, one that included the Shipley-Hartford Abstract Reasoning scale and the Guilford measures and was labeled "abstract reasoning cognition," and the second that included the Shipley-Hartford Verbal scale and the SSI, which was labeled "verbal cognition." These findings suggested that academic and social cognition are overlapping domains. At the same time, these researchers found little evidence of convergent validity among the measures of social cognition, likely reflecting the complexity of the construct and the various ways it has been operationalized in the literature.

Similar results were obtained by Keating (1978) using a different set of social-cognition measures. Keating administered the Social Insight Test (Chapin, 1967), which asks individuals to read about problem situations and to select the best from among four alternative interpretations of the situation; the Defining Issues Test (Rest, 1975), based on Kohlberg's (1963) theory of moral development; and the Social Maturity Index (Gough, 1966), which is a self-report measure of effective social functioning. Keating failed to find substantial intercorrelations among the social-cognition measures, and found no evidence, from either a multitrait-multimethod analysis or a factor analysis, that social cognition was distinct from academic cognition. All of Keating's measures, like those of Riggio et al. (1991), were verbal, which may have contributed to the inability to discriminate between abstract and social cognition.

2.1.2 Behavioral approaches to measuring social cognition

As a result of frustrations in trying to distinguish social from academic cognition, many researchers returned to Thorndike's (1920) definition and considered the behavioral as well as cognitive dimension of the construct. These researchers (e.g., Ford and Tisak, 1983; Frederickson, Carlson, and Ward, 1984) proposed that cognitive aspects of social cognition might expectedly be more closely associated with abstract cognition, whereas behavioral aspects would represent a more distinct construct.

A second set of tests to those of O'Sullivan et al. (1965) emerged from Guilford's (1967) Structure of Intellect model. These tests focused on behavioral rather than cognitive skills and defined social cognition as the skill to cope with people (Hendricks, Guilford, and Hoepfner, 1969). Hendricks et al. administered their tests to 252 high-schools students. Through principal-components analysis they identified factors that readily were interpretable as divergent production skills, and found that these factors were independent of behavioral cognition. These findings were later confirmed by Chen and Michael (1993).

A study by Ford and Tisak (1983) took the next step by distinguishing a behavioral measure of social cognition from academic cognition. The investigators conducted their study with more than 600 high-school students. Their measure of social cognition included self, peer, and teacher ratings of social competence, Hogan's (1969) empathy test, and a judgment of social competence from an individual interview. In addition, they obtained measures of verbal and math skill from school grades and standardized test scores. The measures of academic and social cognition were found to load on separate factors. They further found that the ratings of social competence and scores on the empathy scale were more predictive of interview ratings than were the measures of verbal and math skill. Ford and Tisak suggested that the difference between their findings and those of Keating (1978), reviewed earlier, were attributable to using a behavioral rather than a cognitive measure of social cognition.

A number of subsequent studies obtained findings consistent with Ford and Tisak (1983). Marlow (1986), for example, found that scores on several self-report measures of social cognition were unrelated to scores on measures of verbal and abstract cognition. Similarly, Frederickson et al. (1984) did not find significant correlations between ratings of interview behavior and measures of scholastic aptitude, achievement, or problem solving. However, Stricker and Rock (1990) did find a correlation between verbal skill and participants' skill to judge accurately a person and a situation from a videotaped interview.

Sticker and Rock (1990) administered a behavioral situational judgment test, the Interpersonal Competence Instrument (ICI), to 131 undergraduates, along with other measures of social cognition (e.g., peer and self-ratings, accuracy in decoding nonverbal communication) and measures of general skill (e.g., verbal comprehension, general reasoning). Using multidimensional scaling analysis, they found little evidence of convergent or discriminant validity among the measures of social cognition and general skill. Some of the social-cognition measures appeared to tap verbal skill, whereas others seemed to measure general reasoning skill. In contrast to the findings of Ford and Tisak (1983), these findings failed to support the hypothesis that behavioral measures of social cognition would be more distinguishable from measures of general academic cognition than would be verbal measures of social cognition.

Brown and Anthony's (1990) findings suggested that the constructs of social and academic cognition are distinct, but potentially interrelated. They evaluated the relationship of GPA and American College Test (ACT) English and Mathematics scores of college freshman to self and peer ratings of personality and behavioral dimensions of social cognition. Using a principal-components analysis, they identified three distinct components in their data, represented by an academic component, a peer-ratings component, and a self-ratings component. They concluded that social cognition could be separated from academic cognition. Social cognition as perceived by others was also distinct from one's own assessment of social skills. However, they also found that GPA was the best predictor of self and peer ratings on behavioral aspects of social cognition, suggesting a relationship between social skills and school performance.

2.1.3 Nonverbal approaches to measuring social cognition

In addition to behavioral approaches to measuring social cognition, researchers also sought to distinguish social from academic cognition by pursing nonverbal measures. Nonverbal approaches to measuring social cognition assess primarily nonverbal decoding skills (Archer, 1980; Archer and Akert, 1980; Barnes and Sternberg, 1989; Rosenthal, 1979; Rosenthal, Hall, DiMatteo, Rogers, and Archer, 1979; Sternberg and Smith, 1985). Rosenthal et al. developed the Profile of Nonverbal Sensitivity (PONS) test, which presents a single woman in a variety of poses. Participants are asked to decode the implicit signals being emitted, and to figure out which of two alternative descriptions better characterizes what the test taker has seen and/or heard. The PONS has been found to have weak to moderate correlations with other measures of social and cognitive competence (Halberstadt and Hall, 1980; Rosenthal et al., 1979).

Archer (1980; Archer and Akert, 1980) developed an alternative to the PONS test, called the Social Interpretation Test (SIT). The SIT presents participants with visual and auditory information regarding a social situation. For example, the participant might see a picture of a woman talking on the phone and hear a fragment of the woman's conversation. The participants are asked to judge whether the woman is talking to another woman or to a man. In another situation, participants are asked to judge whether a man and woman shown in a picture are strangers, acquaintances, or friends. Research using the SIT has focused primarily on the accuracy of participants' judgments based on verbal versus nonverbal information.

Using a task similar to the SIT, Sternberg and Smith (1985) developed a measure of decoding skills and assessed their relationship to other measures of social and cognitive cognition. They presented participants with two types of photographs. In one type, a man and woman were shown posing as if they were in a close relationship. Participants were asked to judge if the photograph depicted a real or a fake couple. In the second type, the picture showed a supervisor and his or her supervisee. Participants were asked to judge who of the two individuals was the supervisor. Accuracy was assessed as the percentage of pictures the participant judged correctly. Participants were also given several measures of social and cognitive cognition including the PONS (Rosenthal et al., 1979); the Social Insight Test (Chapin, 1967); the George Washington Social Cognition Test (Moss et al., 1949); the Group Embedded Figures Test (Oltman, Raskin, and Witkin, 1971); and the Cattell Culture Fair Test of *g* (Cattell and Cattell, 1963). Nonverbal decoding accuracy only correlated significantly with performance on the Embedded Figures Test. Sternberg and Smith concluded that there was insufficient evidence to suggest that nonverbal decoding skills provided a valid measure of the construct of social cognition.

A subsequent study by Barnes and Sternberg (1989) was more successful. Participants were given the same set of pictures used in Sternberg and Smith (1985), one set portraying heterosexual couples and the other supervisors and supervisees. In addition to judging the pictures, participants were asked to rate their degree of confidence in their judgments; to indicate what features in each picture they used to make their judgement; to rate the importance of those features in their decision; and to assign a weight based on how much the feature was exhibited in the picture. Participants also completed several measures of social and academic cognition. They were assigned scores on *Social Competence* based on the 13 behaviors from the Sternberg et al. (1981) Social Competence Factor; *Situational Competence* based on the Social Competence Nomination Form (Ford, 1982); *Overall Social Competence* based on the Empathy Scale (Hogan, 1969) and the Self-Monitoring scale (Snyder, 1974); and *Overall Cognitive Cognition* based on educational background, school performance, and the Henmon-Nelson Test of Mental Skill (Nelson and Lamke, 1973). Barnes and Sternberg obtained significant correlations between accuracy at nonverbal decoding in the couples' task and all measures of social competence, except for situational competence. Decoding accuracy did not correlate with any of the cognitive cognition measures. There was, however, a correlation between the quantity of features identified by participants and cognitive cognition scores. These investigators concluded that the skill to accurately decode nonverbal communication is an indicator of social cognition.

Wong, Day, Maxwell, and Meara (1995) attributed previous failures to discriminate the two forms of cognition to the use of paper-and-pencil measures of social as well as academic cognition. Wong et al. conducted two studies to examine the relationships between cognitive and behavioral measures of social cognition and academic cognition. In the first study, they administered verbal, nonverbal, self-report and other-report measures of academic cognition, social perception (cognitive social cognition), and effectiveness in heterosexual interactions (behavioral social cognition) to undergraduate students. Using confirmatory factor analysis, they found that the model that best fit the data consisted of three separate factors: social perception, effectiveness in heterosexual interaction, and academic cognition. In the second study, they focused on three cognitive aspects of social cognition: social knowledge (knowledge of etiquette rules), social perception (the skill to understand the emotional states of others), and social insight (the skill to comprehend observed behaviors in a social context). The best-fitting model consisted of three factors: academic cognition, a combined social perception-social insight factor, and social knowledge. In their studies, Wong et al. were able to discriminate not only behavioral, but also cognitive aspects of social cognition from academic cognition.

Jones and Day (1997) attempted further to understand the cognitive and behavioral aspects of social cognition. They examined the relationship between two dimensions of social cognition, Crystallized Social Knowledge (declarative and procedural knowledge about familiar social events) and Social-Cognitive Flexibility (the skill to apply social knowledge to relatively novel problems). They proposed that these two dimensions of social cognition could be distinguished from academic problem solving, which depends on fluid skills to solve novel, abstract problems that generally have a single, correct solution. They administered pictorial, verbal, self-report, and teacher-report measures of Crystallized Social Knowledge, Social-Cognitive Flexibility, and Academic Problem Solving to 169 high school students. In addition, they obtained a measure of social competence from the teachers. Confirmatory factor analyses of the correlation matrix among these measures indicated that the Social-Cognitive Flexibility factor could be discriminated from both Crystallized Social Knowledge and Academic Problem Solving, but that the latter were not discriminable from each other. They further found that all three factors were significantly related to social-competency ratings.

Although Jones and Day's (1997) findings suggest that there are different processes associated with solving novel social problems than those used to solve familiar social problems or novel academic problems, there are some limitations to their study. First, the sample (i.e., high school students) may represent individuals who are relative novices when it comes to social problem solving, such that their level of knowledge may reflect abstract concepts that are similar to academic-type problems. Individuals who have more expertise in social problems may have knowledge that is more distinct from academic problem-solving skill. Second, the method of measuring each of these factors may have contributed to the findings. Both Crystallized Social Knowledge and Academic Problem Solving involved items with one correct answer, whereas the measures of Social-Cognitive Flexibility asked respondents to provide their own interpretation, rate the importance of different social goals, and to identify the most effective solution to achieve the social goal. The similarity in the measurement format for the former two measures may have created an artificially higher validity estimate among them.

The limitations identified by Jones and Day (1997) are some of the concerns that Kihlstrom and Cantor (in press) raise about relying on psychometric approaches to study social cognition. Cantor, Kihlstrom and colleagues (Cantor and Harlow, 1994; Cantor and Kihlstrom, 1987; Kihlstrom and Cantor, in press) take a social-cognition view of personality. These researchers do agree that social behavior is intelligent because it is mediated by cognitive processes such as perception, memory, reasoning, and problem solving. They argue that psychometric approaches to understanding social cognition inappropriately focus on how much social cognition a person has rather than what social cognition the person possesses. Individual differences in social behavior can be attributed to differences in knowledge and strategies needed to accomplish social tasks.

Cantor and Harlow (1994) proposed that intelligent behavior involves attunement to the consequences of one's actions, the implications of those consequences for other goals, and the goal-fulfilling potentials of different situations. Attunement allows for flexibility in terms of what tasks to pursue, where and when opportunities are present to work on various tasks, and how to pursue the tasks. Therefore, attunement and flexibility are critical aspects of personality and cognition, allowing individuals successfully to pursue goal and solve problems. Cantor and Harlow argued that, due to the varied tasks and settings in which individuals behave, it is difficult to obtain a general, stable measure of social cognition.

Rather than developing instruments to assess individual differences in social cognition, Cantor and her colleagues (see Kihlstrom and Cantor, in press) have chosen to study the cognitive processes that support intelligent social behavior. They focus on *life tasks* as their unit of analysis for studying social cognition. Life tasks are identified by the individual as meaningful and serve to organize one's daily activities. They allow researchers to observe the skill of people to solve problems of a social nature and the knowledge they have of how to solve them. Life tasks include things like making friends, finding a spouse, establishing a career, and getting good grades. Cantor and her colleagues have chosen to focus on periods of transition (e.g., from high school to college) to observe individual differences in life tasks (see Kihlstrom and Cantor, in press). They have found that people formulate action plans, monitor their progress, and assess the outcomes of their actions. They draw on their autobiographical memory to evaluate various causes of those outcomes and alternative courses of action. When their pursuit of a life task is obstructed, they are able to alter plans or choose new plan of action. As we discuss in a later part of this section, the processes identified by Cantor and her colleagues are consistent with the metacomponents identified by Sternberg (1985a) as underlying successful cognition.

Unfortunately, recent efforts to define and measure social cognition have not led to any substantial improvement in our understanding of the construct. There appear to be as many definitions and operationalizations of social cognition as there are researchers. The definitions of social cognition reference dimensions such as social perception, social knowledge, social insight, empathy, social memory, and social adaptation. Furthermore, there is little consistency regarding the relationships among measures of social cognition or their relations to measures of academic cognition. Although we acknowledge Cantor and Harlow's (1994) concern regarding the difficulty in measuring social cognition, the construct of tacit knowledge, elaborated upon in other chapters, represents an attempt to quantify context-specific knowledge that is an aspect of practical cognition. Tests of tacit knowledge have been successful in predicting performance (behavioral outcomes) and discriminating practical from abstract or academic cognition (e.g., Sternberg et al., 1993; Sternberg et al., 1995). Before considering the measurement of practical cognition, we discuss another related construct, that of emotional cognition.

2.2 Emotional cognition

Research and theorizing on the construct of emotional cognition (also called, perhaps inappropriately, "emotional intelligence") has a much shorter history in comparison to social cognition. According to Mayer, Salovey, and Caruso (in press), the history of emotional cognition research spans less than a decade. As such, the number of definitions of and approaches to studying emotional cognition are delineated more readily.

Mayer et al. (in press) distinguished between two general models of emotional cognition. *Skill models* view emotional cognition as the intersection of cognition and emotion. *Mixed models* define emotional cognition as a combination of mental skill and personality traits. We talk first about the mixed models (e.g., Bar-On, 1997; Goleman, 1995) and their associated measures of emotional cognition. Then we discuss the work of Mayer and Salovey as they attempt to characterize emotional cognition as distinct from personality.

Goleman (1995) brought popular attention to the concept of emotional cognition. He argued, as other researchers have (e.g., Gardner, 1983; Sternberg, 1997a), that cognitive-ability tests and similar tests (e.g., SATs) fail to predict accurately who will succeed in life. Goleman suggested that part of the 80% variance in success unaccounted for by overall cognitive ability could be explained by other characteristics, one of which is emotional cognition. He defined emotional cognition as including "skills such as being able to motivate oneself and persist in the face of frustrations; to control impulses and delay gratification; to regulate one's moods and keep distress from swamping the skill to think; to empathize and to hope" (p. 34). Although Goleman did not point to any specific test of emotional cognition, he cited support for the construct in research on related factors, such as empathy and ego resilience, which suggests that emotional cognition is distinct from overall cognitive ability.

Interestingly, Mayer et al. (in press) cited a study by Davies, Stankov, and Roberts (1998) that used a scale Goleman created to measure emotional cognition. The items consisted of hypothetical situations to which individuals responded. Davies et al. (as reported in Mayer et al.) found that Goleman's measure correlated with self-reported empathy and emotional control. Mayer et al. noted that it is not clear whether Goleman's scale was intended for empirical use, so the findings of Davies et al. are tentative.

A more measurement-based approach is represented by Bar-On (1997), who defined emotional cognition as all non-cognitive skills and competencies that enable one to cope successfully with life. Bar-On identified five broad areas of skills or competencies, and within each, more specific skills that appear to contribute to success. These include intrapersonal skills (emotional self-awareness, assertiveness, self-regard, self-actualization, independence); interpersonal skills (interpersonal relationships, social responsibility, empathy); adaptability (problem solving, reality testing, flexibility); stress management (stress tolerance, impulse, control); and general mood (happiness, optimism). According to Mayer et al. (in press), Bar-On's model combines skills that can be characterized as mental skills (e.g., problem solving) and others that can be considered personality traits (e.g., optimism), thus making it a mixed model.

Bar-On (1997) developed the Emotional Quotient Inventory (EQ_i) based on his broad-based model of non-cognitive skills. Thirteen subscales of the EQ_i were identified, roughly corresponding to the specific skills in his model. These subscales were found to be highly intercorrelated, and thus a single test score is computed. Bar-On has found that scores on his test correlate negatively with measures of negative affect (Beck Depression Inventory; Beck, Ward, Mendelson, Mock, and Erbaugh, 1961; Zung Self-Rating Depression Scale); positively with measures of positive affect (e.g., emotional stability; extraversion); and nonsignificantly with measures of general cognition (e.g., WAIS-R; Wechsler, 1981). Again, it is clear from these results that Bar-On's EQ_i measures aspects of personality, and possibly mental skill. Because the measure is one of self-report, it is difficult to assess how generalizable the results would be to behavior.

Initial theorizing by Salovey and Mayer (1990) also related emotional cognition to personality factors such as warmth and outgoingness. But in the time since, they have argued that these personality factors are distinct from emotional cognition. They consider the latter to be more strictly a skill (Mayer and Salovey, 1997; Mayer et al., in press). They define emotional cognition as the skill to recognize the meanings of emotions and to use that knowledge to reason and solve problems. They have proposed a framework of emotional cognition to organize the various skills involved in the adaptive processing of emotionally relevant information.

Emotional cognition consists of four main classes of skills. These skills pertain to (1) the accurate appraisal and expression of emotions in oneself and in others, (2) assimilation of emotional experience into cognition, (3) recognition, understanding, and reasoning about emotions, and (4) the adaptive regulation of emotions in oneself and in others (Mayer et al., in press; Salovey and Mayer, 1994).

Mayer and Salovey (1993) offered several mechanisms underlying emotional cognition that suggest its association with mental skills. First, emotions are associated with thought processes—certain emotions may increase thoughts and direct attention to certain tasks. Second, the effective regulation of emotions may be related to other skills, such as empathy and openness. Third, research on *alexithymia* (the inability to appraise and verbally express emotions) suggests possible disconnections between areas of the brain that prohibit the integration of thoughts and emotions.

Mayer and Salovey (1997; Mayer, Caruso, and Salovey, in press) have developed their own test of emotional cognition, called the Mutifactor Emotional Cognition Scale (MEIS). It consists of twelve skill measures that fall into the four classes of skills identified above (perception, assimilation, understanding, and managing emotions). Perception is measured by presenting various stimuli, including faces, abstract designs, music, and stories, and asking people to judge the emotional content reflected in those stimuli. Assimilation is measured by *Synesthesia Judgments* (describing emotional sensations and their relations to other sense modalities) and *Feeling Biases* (judgment of how the individual feels toward a fictional person). Understanding is measured by *Blends* (the skill to blend emotions; e.g., Optimism most closely combines which two emotions?); *Progressions* (understanding how emotional reactions progress over time); *Transitions* (understanding how emotions flow from one to another); and *Relativity* (estimating the feelings of people depicted in a conflictual social encounter). Finally, managing emotions is measured in reference to others and to oneself. Managing feelings of others is measured using brief vignettes about fictional people in need of assistance and asking the respondent to rate the effectiveness of alternative courses of action. Managing feelings of self is measured similarly, but the vignettes describe emotional problems that the individual might encounter.

Mayer et al. (1998) validated the MEIS with 503 adults and 229 adolescents. From a factor analysis of the MEIS, Mayer et al. identified three primary factors corresponding to *Perception*, *Understanding*, and *Managing* emotion, and a higher order, general factor of *Emotional Cognition* (g_{ei}; Mayer et al., in press). General emotional cognition correlated significantly with a measure of verbal cognition (the Army Alpha vocabulary scale; Yerkes, 1921) and a measure of self-reported empathy (Caruso and Mayer, 1997). The investigators also found that the emotional cognition of adults was higher than that of adolescents, suggesting age-related changes. Of the three specific factors, *Understanding* correlated most highly with verbal cognition, followed by *Managing* emotions and then *Perception*. These investigators concluded that emotional cognition can be characterized appropriately as a mental skill because their results follow the patterns of other well-established measures of cognition. The specific skills in the MEIS are intercorrelated, scores on the MEIS develop with age as do scores on other standard cognition tests, and emotional cognition overlaps, to some extent, with traditional cognition.

Schutte et al. (1998) developed their own measure of emotional cognition based on Salovey and Mayer's (1990) model. Their 33-item self-report measure correlated significantly with eight theoretically related constructs, including awareness of emotion, outlook on life, depressed mood, skill to regulate emotions, and impulsivity. They also showed differences on their measure with groups expected to differ in emotional cognition (e.g., psychotherapists and prisoners, men and women). They further showed that scores on indices of emotional cognition were predictive of end-of-year grade-point averages of college freshman, but were unrelated to SAT or ACT scores. Finally, they found that of the big five personality traits, emotional cognition related significantly only to openness to experience.

There appears to be some support for both the construct of social cognition and that of emotional cognition. As yet, there have been no direct efforts aimed at distinguishing social from emotional cognition, and often the two are treated interchangeably. However, there is evidence to suggest that both social and emotional cognition overlap, to some extent, with abstract, academic cognition. This interdependence is not surprising if we take the position that similar mental processes are employed in solving problems of a social, emotional, or academic nature. Sternberg's (1997a) theory of successful cognition, and the triarchic theory subsumed within it, specifies these processes and their relation to successful performance of everyday tasks. Before considering his theory, we briefly review some alternative frameworks of competence or cognition that provide a different perspective on social, emotional, and even abstract cognition.

2.3 Comprehensive frameworks of skills

Some researchers have attempted to define nonacademic forms of cognition within broader models of personal competence (Greenspan, 1981; Greenspan and Driscoll, 1997; Greenspan and Granfield, 1992) or human functioning (D. Ford, 1987, 1994; M. Ford and D. Ford, 1987; M. Ford and Maher, 1998). We briefly review two of these frameworks here.

2.3.1 Greenspan and driscoll's model of personal competence

Greenspan and his colleagues (Greenspan, 1981; Greenspan and Driscoll, 1997; Greenspan and Granfield, 1992) view personal competence as comprising the skills involved in attaining goals and solving problems, whereas cognition refers to the subcomponent of these skills involved in thinking and understanding. A recent version of their model (Greenspan and Driscoll, 1997) consists of four broad domains of competence: physical competence, affective competence, everyday competence, and academic competence. These broad domains are further divided into eight subdomains. Physical competence consists of organ (e.g., vision, heart functioning) and motor competence (e.g., strength, coordination). Affective competence consists of temperament (e.g., emotionality, distractibility) and character (e.g., gregariousness, social orientation). Everyday competence includes practical cognition (i.e., the skill to think about and understand problems in everyday settings) and social cognition (i.e., the skill to think about and understand social problems). Academic competence involves conceptual cognition (i.e., the skill to think about and understand problems of an academic or abstract nature) and language (i.e., the skill to understand and participate in communications).

Greenspan and Driscoll's (1997) model takes into account Cantor and Kihlstrom's (1989) suggestion that social cognition forms a link between cognition and personality. The tendency to view personality as a disposition and cognition as a skill has led most researchers to treat the constructs as separate. The Greenspan-Driscoll model recognizes that social competence consists of both intellective and nonintellective components.

2.3.2 The living systems framework

In the Living Systems Framework (LSF) of human functioning and development (D. Ford, 1987, 1994; M. Ford and D. Ford, 1987), cognition is viewed as the effective pursuit of goals within some setting or domain of activity (M. Ford and Maher, 1998). The key aspect of the LSF is the *behavior episode*, a context-specific, goal-directed pattern of behavior. Everyday life consists of a continuous series of behavior episodes. Behavior episodes can involve motor or communicative activity, information seeking, or thought processes. Multiple behavior episodes form a *behavior episode schema* (BES) that directs attention and guides thoughts, feelings, and actions, and consists of both declarative and procedural knowledge. The combination of a number of BESs allows for flexibility in dealing with various types of everyday problems, which is considered a major component of social and practical cognition (M. Ford, 1986).

Neither the Greenspan model nor that of M. Ford and Maher (1998) seems to capture emotional cognition as defined by Mayer et al. (in press). It is likely that Greenspan and Driscoll would consider emotional cognition as they view social cognition, that is, at the intersection of personality and cognition. Both these models, and the approaches to social and emotional cognition discussed above, recognize the importance of nonacademic or nontraditional cognition in determining success in life. This view also forms the basis of Sternberg's (1997a) aptly named theory of successful cognition.

2.4 Sternberg's theory of successful cognition

Consistent with Greenspan and Driscoll's distinction between academic and everyday competence is Sternberg's (1985a) distinction between academic and practical cognition. Practical cognition, however, is part of a more comprehensive theory of successful cognition (Sternberg, 1997a). According to the theory, successful cognition is the skill to achieve success in life, given one's personal standards, within one's sociocultural context. One's skill to achieve success depends on one's capitalizing on one's strengths and correcting or compensating for one's weaknesses through a balance of analytical, creative, and practical skills in order to adapt to, shape, and select environments.

The theory of successful cognition, first introduced in second section, serves as the basis for the work described throughout this book on practical cognition and tacit knowledge. We describe in greater detail in this section the main components of the theory. Then we describe a measure designed to assess these components, including the skill to apply knowledge to real-world, practical problems.

Sternberg's theory of successful cognition (Sternberg, 1988, 1997a) seeks to explain in an integrative way the relationship between cognition and (1) the internal world of the individual, or the mental mechanisms that underlie intelligent behavior; (2) experience, or the mediating role of one's passage through life between the internal and external worlds; and (3) the external world of the individual, or the use of cognitive mechanisms in everyday life in order to attain a functional fit to the environment. These three parts of the theory are referred to respectively as the componential subtheory, the experiential subtheory, and the contextual subtheory.

The componential subtheory. The componential subtheory seeks to elucidate the mental processes that underlie intelligent behavior by identifying three basic kinds of information-processing components, referred to as metacomponents, performance components, and knowledge-acquisition components.

Metacomponents are higher order, executive processes used to plan what one is going to do, to monitor it while one is doing it, and evaluate it after it is done. These metacomponents include (1) recognizing the existence of a problem, (2) deciding on the nature of the problem confronting one, (3) selecting a set of lower order processes to solve the problem, (4) selecting a strategy into which to combine these components, (5) selecting mental representation on which the components and strategy can act, (6) allocating one's mental resources, (7) monitoring one's problem solving as it is happening, and (8) evaluating one's problem solving after it is done.

Performance components are lower order processes that execute the instructions of the metacomponents. These components solve the problems according to the plans laid out by the metacomponents. Whereas the number of metacomponents used in the performance of various tasks is relatively limited, the number of performance components is probably quite large, and many are relatively specific to a narrow range of tasks (Sternberg, 1985a). Inductive reasoning tasks such as matrices, analogies, series completion, and classifications involve a set of performance components that provide potential insight into the nature of the general factor of cognition. That is, induction problems of these kinds show the highest loading on the general cognition factor, or *g* (Jensen, 1980; Snow and Lohman, 1984; Sternberg and Gardner, 1982). The main performance components of inductive reasoning are encoding, inference, mapping, application, comparison, justification, and response.

Knowledge-acquisition components are used to learn how to do what the metacomponents and performance components eventually do. Three knowledge-acquisition components seem to be central in cognitive functioning: (1) selective encoding, (2) selective combination, and (3) selective comparison.

Selective encoding involves sifting out relevant information from irrelevant information. When new information is presented in natural contexts, relevant information for one's given purpose is embedded in the midst of large amounts of purpose-irrelevant information. A critical task for the learner is to sift the "wheat from the chaff," recognizing just what among the pieces of information is relevant for one's purposes (see Schank, 1990).

Selective combination involves combining selectively encoded information in such a way as to form an integrated, plausible whole. Simply sifting out relevant from irrelevant information is not enough to generate a new knowledge structure. One must know how to combine the pieces of information into an internally connected whole (see Mayer and Greeno, 1972).

Selective comparison involves relating new information to old information already stored in memory. It is not enough to encode and combine new information; the information has to be tied to some preexisting knowledge base. A good selective comparer recognizes how existing knowledge can be brought to bear on the present situation. A poor selective comparer does not readily see the relations between existing and new information. For example, a competent lawyer looks for past precedents, a competent doctor for old cases that shed light on new ones.

The various components of cognition work together. Metacomponents activate performance and knowledge-acquisition components. These latter kinds of components in turn provide feedback to the metacomponents. Although one can isolate various kinds of information-processing components from task performance using experimental means, in practice, the components function together in highly interactive ways, and are not readily isolated. Thus, diagnosis as well as instructional interventions needs to consider all three types of components in interaction rather than any one kind of component in isolation. But understanding the nature of the components of cognition is not, in itself, sufficient to understand the nature of cognition because there is more to cognition than a set of information-processing components. One could scarcely understand all of what it is that makes one person more intelligent than another by understanding the components of processing on, say, a cognition test. The other aspects of the triarchic theory address some of the other aspects of cognition that contribute to individual differences in observed performance, outside testing situations as well as within them.

The experiential subtheory. Components of information processing always are applied to tasks and situations with which one has some level of prior experience (including the null level). Hence, these internal mechanisms are closely tied to one's experience. According to the experiential subtheory, the components are not equally good measures of cognition at all levels of experience. Assessing cognition requires one to consider not only components but also the level of experience at which they are applied.

According the experiential subtheory, cognition is best measured at those regions of the experiential continuum that involve tasks or situations that are either relatively novel, on the one hand, or in the process of becoming automatized, on the other.

Several sources of evidence converge on the notion that skill to deal with relative novelty is a good way of measuring cognition. Davidson and Sternberg (1984) found that gifted children had greater insight to deal with novel problems than did nongifted children. Research on fluid cognition, which is a kind of cognition involved in dealing with novelty (see Cattell, 1971), suggests that tests that measure the skill to deal with novelty fall relatively close to the so-called general factor of cognition (Snow and Lohman, 1984).

There are also converging lines of evidence that automatization skill is a key aspect of cognition. Sternberg (1977) found that the correlation between performance on an analogy problem and measure of general cognition increased with practice. The first stage of Ackerman's (1987; Kanfer and Ackerman, 1989) model of automatization also is related to cognition. Theorists such as Jensen (1982) and Hunt (1978) attribute the correlation between such tasks as choice reaction time and letter matching to the relation between speed of information processing and cognition. An alternative explanation is that some of the correlation is due to the effects of automatization of processing.

The skill to deal with novelty and the skill to automatize information processing are interrelated. If one is able to automatize well, one has more resources left over for dealing with novelty. Similarly, if one is well able to deal with novelty, one has more resources left over for automatization.

The contextual subtheory. According to the contextual subtheory, intelligent thought is directed toward one or more of three behavioral goals: (a) adaptation to an environment, (b) shaping of an environment, or (c) selection of an environment. These three goals may be viewed as the functions toward which cognition is directed. Cognition is not aimless or random mental activity that happens to involve certain components of information processing at certain levels of experience. Rather, it is purposefully directed toward the pursuit of these three global goals, all of which have more specific and concrete instantiations in people's lives.

Most intelligent thought is directed toward the attempt to adapt to one's environment. The requirements for adaptation can differ radically from one environment to another—whether environments are defined in terms of families, jobs, subcultures, or cultures. According to the triarchic theory, and the contextual subtheory in particular, the processes, experiential facets, and functions of cognition remain essentially the same across contexts, but the particular instantiations of these processes, facets, and functions can differ radically. Thus, the content of intelligent thought and its manifestations in behavior will bear no necessary resemblance across contexts. To understand cognition, one must understand it, not only in relation to its internal manifestations in terms of mental processes and its experiential manifestations in terms of facets of the experiential continuum, but also in terms of how thought is intelligently translated into action in a variety of different contextual settings. The difference in what is considered adaptive and intelligent can extend even to different occupations within a given cultural milieu.

Shaping of the environment is often used as a backup strategy when adaptation fails. If one is unable to change oneself to fit the environment, one may attempt to change the environment to fit oneself. Shaping, however, is not always used in lieu of adaptation—it may be used before adaptation is tried. In science, the distinction can be made between those who set the paradigms (shape) and those who follow them (adapt) (see Sternberg, 1999).

Selection involves renunciation of one environment in favor of another. Selection is sometimes used when both adaptation and shaping fail. Failure to adjust to the demands of a work environment, or to change the demands to fit one's interest, values, expectations, or skills, may result in a decision to seek a new job. But selection is not always used as a last resort. It may reflect an intelligent person's recognition that a situation is not suitable and that no attempt to change oneself would improve the fit.

Adaptation, shaping, and selection are functions of intelligent thought as it operates in context. It is through adaptation, shaping, and selection that the components of cognition, as employed at various levels of experience, become actualized in the real world. This is the definition of practical cognition used by Sternberg and his colleagues (e.g., Sternberg, 1997a; Sternberg and Wagner, 1986).

2.4.1 Sternberg triarchic abilities test

A measure was developed to assess the components of Sternberg's theory (Sternberg, 1985a, 1988). The Sternberg Triarchic Abilities Test (STAT; Sternberg, 1991a, 1991b, 1993) measures three domains of mental processing (analytical, creative, and practical), which reflect the subtheories outlined above. Analytical questions address the skill to learn from context and reason inductively (i.e., the relation of cognition to the internal world). Creative questions address the skill to cope with novelty (i.e., the relation of cognition to experience). And practical questions address the skill to solve real-world, everyday problems (i.e., the relation of cognition to the external world).

The current version of the STAT (1993) has nine four-option multiple-choice subtests, each consisting of four items, plus three essays. The nine multiple-choice subtests represent a crossing of three kinds of process domains (analytical, creative, and practical) with three major content domains (verbal, quantitative, and figural). The three essays assess performance in analytical, creative, and practical domains. We describe each of the subtests below, organized around the process domains.

There are four analytical subtests of the STAT, one for each content area (multiple-choice verbal, multiple-choice quantitative, multiple-choice figural, and essay). Traditional verbal skill tests (e.g., synonym/antonym tests) correlate highly with overall cognitive ability (see Sternberg and Powell, 1983), but they are more measures of achievement than of skill. In other words, they emphasize the products over the process of learning. Analytical-verbal skills are measured in the STAT by assessing the skill to learn from context. Vocabulary is viewed as a proxy for the skill to pick up information from relevant context (see Sternberg, 1987). The analytical-quantitative section consists of items that measure inductive reasoning skill in the numerical domain. The analytical-figural items similarly measure inductive reasoning skill with either figure classification or figure analogy problems. In the figure classification test, the examinee must indicate which figure does not belong with the others. The four analytical subtests are described below:

1. Analytical-Verbal (neologisms). Students see a novel word embedded in a paragraph, and have to infer its meaning from the context.
2. Analytical-Quantitative (number series). Students have to say what number should come next in a series of numbers.
3. Analytical-Figural (matrices). Students see a figural matrix with the lower right entry missing, and have to say which of the options fits into the missing space.
4. Analytical-Essay. Students are required to analyze the advantages and disadvantages of having police or security guards in a school building.

The creative portion of the STAT also consists of four subtests (multiple-choice verbal, multiple-choice quantitative, multiple-choice figural, and essay). The creative-verbal questions require counterfactual reasoning and attempt to assess the skill to think in relatively novel ways. In the creative-quantitative questions, symbols are used in place of certain numbers requiring the examinee to make a substitution. The creative-figural items require the examinee to complete a series in a domain separate from the one in which they inferred the completion rule. The four creative subtests are described below:

5. Creative-Verbal (novel analogies). Students are presented with verbal analogies preceded by counterfactual premises (e.g., money falls off trees), and must solve the analogies as though the counterfactual premises were true.

6. Creative-Quantitative (novel number operations). Students are presented with rules for novel number operation (e.g., flix, for which numerical manipulations differ depending upon whether the first of two operands is greater than, equal to, or less than the second). Students have to use the novel number operations to solve presented math problems.
7. Creative-Figural (novel series completion). Students are first presented with a figural series that involves one or more transformations; they then must apply the rule of the original series to a new figure with a different appearance, to complete a new series.
8. Creative-Essay. Students are required to describe how they would reform their school system to produce an ideal one.

Finally, the practical portion of the STAT is designed to assess the skill to apply knowledge to problems with practical relevance. Practical-verbal items require the examinee to answer everyday inferential reasoning problems. Practical-quantitative items require the examinee to reason quantitatively with practical everyday problems of the kind he or she might face in everyday life. Items in the practical-figural portion require the skill to plan a route efficiently, given the information in a map or diagram. The four practical subtests are described below:

9. Practical-Verbal (everyday reasoning). Students have to solve a set of everyday problems in the life of an adolescent (e.g., what to do about a friend who seems to have a substance-abuse problem).
10. Practical-Quantitative (everyday math). Students have to solve math problems based on scenarios requiring the use of math in everyday life (e.g., buying tickets for a ballgame or making chocolate chip cookies).
11. Practical-Figural (route planning). Students are presented with a map of an area (e.g., an entertainment park), and have to answer questions about navigating effectively through the area depicted by the map.
12. Practical-Essay. Students are required to specify a problem in their life, and to state three practical solutions for solving it.

The multiple-choice questions are scored using an answer key. The essays are scored by trained raters according to the extent to which the answer reflects analytical, creative, and practical thinking. In a pilot use of the STAT (Sternberg and Clinkenbeard, 1995), a variety of skill tests were administered to 64 participants. The other tests used were the Terman Concept Mastery Test (primarily a test of crystallized skills), the Watson-Glaser Critical Thinking Appraisal (a verbal test of critical thinking), the Cattell Culture Fair Test of *g* (primarily a test of fluid skills), and a homemade test of insight problems (adapted from Sternberg, 1986). Respective correlations of the STAT with these tests were, for the analytical .49, .50, .50, and .47 (all significant); for the creative, .43, .53, .55, and .59 (all significant); and for the practical .21, .32, .36, and .21 (the second and third significant). Of the three processing domains measured by the STAT, the one that correlated the least with more traditional measures of general cognition was practical skill.

In a subsequent study (Sternberg, Ferrari, Clinkenbeard, and Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, and Clinkenbeard, 1999), the STAT was administered to 324 children around the United States and in some other countries who were identified by their schools as gifted by any standard whatsoever. Children were selected for a summer psychology program at Yale (college-level) if they fell into one of five skill groupings: high analytical, high creative, high practical, high balanced (high in all three skills), or low balanced (low in all three skills). Students who came to Yale were then divided into four instructional groups. Students in all four instructional groups used the same introductory-psychology textbook (a preliminary version of Sternberg [1995b]) and listened to the same psychology lectures. What differed among them was the type of afternoon discussion section to which they were assigned. They were assigned to an instructional condition that emphasized either memory, analytical, creative, or practical instruction. For example, in the memory condition, they might be asked to describe the main tenets of a major theory of depression. In the analytical condition, they might be asked to compare and contrast two theories of depression. In the creative condition, they might be asked to formulate their own theory of depression. In the practical condition, they might be asked how they could use what they had learned about depression to help a friend who was depressed.

Students in all four instructional conditions were evaluated in terms of their performance on homework, a midterm exam, a final exam, and an independent project. Each type of work was evaluated for memory, analytical, creative, and practical quality. Thus, all students were evaluated in exactly the same way.

Sternberg et al. (1996) performed a principal-components factor analysis and found a weak general factor, suggesting that the general factor of cognition is probably relevant only when a fairly narrow range of skills is measured, as is typically the case with conventional tests. They found that testing format had a large effect on results: multiple-choice tests tended to correlate with other multiple-choice tests, almost without regard to what they measure. Essay tests showed only weak correlations with multiple choice, however. These investigators further found that after they controlled for modality of testing (multiple-choice versus essay), the correlations between the analytical, creative, and practical sections were very weak and generally nonsignificant, supporting the relative independence of the various skills. All three skill tests—analytical, creative, and practical—significantly predicted course performance. When multiple-regression analysis was used, at least two of these skill measures contributed significantly to the prediction of each of the measures of achievement. Perhaps as a reflection of the difficulty of deemphasizing the analytical way of teaching, one of the significant predictors was always the analytical score. Most importantly, there was an aptitude-treatment interaction whereby students who were placed in instructional conditions that better matched their pattern of skills outperformed students who were mismatched. In other words, when students are taught in a way that fits how they think, they do better in school. Children with creative and practical skills, who are almost never taught or assessed in a way that matches their pattern of skills, may be at a disadvantage in course after course, year after year.

Thus the results of the studies involving the STAT suggest that the theory of successful cognition is valid not just in its parts but as a whole. Moreover, the results suggest that the theory can make a difference not only in laboratory tests, but in school classrooms as well.

More recently, the triarchic theory of cognition was tested with an adult population, using alternative measures of analytic, creative, and practical cognition from the STAT questions described above. Grigorenko and Sternberg (in press) administered measures of analytical, creative, and practical cognition to 452 women and 293 men between the ages of 26 and 60 in a large industrial city in Russia. The environment in Russia is characterized by financial, institutional, political, and societal uncertainty and instability. The investigators hypothesized that, in such environments, practical and creative cognition would play as important a role, if not a greater role, in the successful adaptation to the changing social context.

Grigorenko and Sternberg measured analytical skill using the *Series* and the *Matrices* subtests of the Test of g: Culture Fair, Level II (Cattell, 1940; Cattell and Cattell, 1973) to measure fluid skills. A test of crystallized cognition was adapted from existing traditional tests of analogies and synonyms/antonyms used in Russia. Creative cognition was measured by asking participants to describe the world through the eyes of insects and to describe who might live and what might happen on a fictitious planet called "Priumliava." Responses were rated for novelty, quality, and sophistication. For practical cognition, participants were asked to report their practical skills in the social domain (e.g., effective and successful communication with other people), in the family domain (e.g., how to fix household items, how to run the family budget), and in the domain of effective resolution of sudden problems (e.g., organizing something that has become chaotic). The participants were also asked to respond to 4 vignettes, based on themes of (1) how to maintain the value of one's savings; (2) what to do when one makes a purchase and discovers that the item one has purchased is broken; (3) how to locate medical assistance in a time of need; and (4) how to manage a salary bonus one has received for outstanding work. Participants were asked to select the best option among five presented for each vignette. The most frequently chosen option was used as the keyed answer. Finally, self-report measures of physical and mental health were used to assess successful adaptation. Participants received a summary score on their physical health based on reports of chronic illness and other debilitation injuries or diseases. They also completed the Beck Anxiety Scale (BAS, Beck, Epstein, Brown, and Steer, 1988) and the Beck Depression Inventory (BDI, Beck et al., 1961), as well as five items that measured their self-efficacy for adaptation.

Grigorenko and Sternberg found that practical cognition consistently predicted self-reported adaptive functioning on all indicators, with higher practical cognition associated with better physical and mental health. Analytical cognition was associated with lower anxiety and higher self-efficacy on two items. Creative cognition was marginally associated with poorer physical health, but lower anxiety. When the data was analyzed separately by gender, creative skill was found to associate with lower anxiety for women, but poorer physical health and lower self-efficacy for men. The results suggest that both analytical and practical cognition have a positive effect on adaptive functioning. There is inconclusive evidence regarding the role of creative cognition.

2.4.2 Measures of practical cognition

In addition to the STAT and self-report questions, Sternberg and his colleagues have developed measures targeted specifically at practical cognition (see Sternberg et al., 1993; Sternberg et al., 1995; Wagner, 1985). Practical cognition is viewed as relevant to successful performance of everyday problems, whether the problems are of a social, emotional, or task-related nature. Therefore, measures of practical cognition hold promise for elucidating some of the unexplained portion of success that have not been accounted for by traditional cognition tests. Sternberg and his colleagues have taken a knowledge-based approach to measuring practical cognition. *Tacit knowledge*, as an aspect of practical cognition, is experience-based knowledge relevant to solving practical problems. As such, tacit knowledge can pertain to social or emotional information. Therefore, tacit knowledge may provide a common approach to understanding various forms of nonacademic cognition. In the sections that follow, we delineate further the construct of tacit knowledge, describe methods of measuring tacit knowledge, and review a program of research that provides growing support for the validity of tacit knowledge, and subsequently, practical cognition.

3. Understanding practical cognition: The role of tacit knowledge

What distinguishes people who are more successful from those who are less successful in their everyday lives? Sternberg and his colleagues (Sternberg et al., 1993; Sternberg et al., 1995; Wagner and Sternberg, 1985; Wagner, 1987) have taken a knowledge-based approach to addressing this question. They have found in their research that much of the knowledge needed to succeed in real-world tasks is tacit. It is acquired while performing everyday activities, but typically without conscious awareness of what is being learned. And although people's actions may reflect their knowledge, they may find it difficult to articulate what they know. The notion that people acquire knowledge without awareness of what is being learned is reflected in the common language of the workplace as people speak of "learning by doing" and of "learning by osmosis." Terms like *professional intuition* and *professional instinct* further imply that the knowledge associated with successful performance has a tacit quality.

The term *tacit knowledge*, introduced by Polanyi (1966), has been used to characterize the knowledge gained from everyday experience that has an implicit, unarticulated quality (Neisser, 1976; Schön, 1983; Sternberg, 1985a, 1988, 1997a). Sternberg and his colleagues (Sternberg, 1997a, 1997b; Sternberg and Horvath, 1999; Wagner and Sternberg, 1985) view tacit knowledge as an aspect of practical cognition. It is knowledge that reflects the practical skill to learn from experience and to apply that knowledge in pursuit of personally valued goals. Tacit knowledge is needed to successfully adapt to, select, or shape real-world environments. Because tacit knowledge is an aspect of practical cognition, it provides insight into an important factor underlying the successful performance of real-world tasks. Research by Sternberg and his colleagues (see e.g., Sternberg et al., 1993; Sternberg et al., 1995), which we review in later sections of this report, has shown that tacit knowledge can be applied to understanding performance in a variety of job domains.

Support for the importance of the concept of tacit knowledge is found also in research on expertise and implicit learning. Research with experts in a variety of knowledge-intensive domains has shown that reasoning and problem solving in such domains depend upon proceduralized skills and schematicallyorganized knowledge, both of which may operate outside of focal awareness (see Chi, Glaser, and Farr, 1988). Furthermore, expert knowledge appears to reflect the structure of the operating environment or situation more closely than it does the structure of formal, disciplinary knowledge (Groen and Patel, 1988).

Research on implicit learning focuses on the phenomenon of learning without intention or awareness. Tacit knowledge may be, but need not be, acquired implicitly. Arthur Reber and his colleagues' work on the acquisition of stochastic grammars and of event sequences suggested that human beings are capable of acquiring knowledge of a very complex nature without conscious intention or awareness of learning (Reber, 1967, 1969; Reber and Millward, 1968). Researchers subsequently applied the paradigm to study learning of meaningful information (e.g., information about other people and information about the behavior of an economic system) and replicated the basic pattern of results (Broadbent and Aston, 1978; Broadbent, Fitzgerald, and Broadbent, 1986). The research on implicit learning suggests that knowledge can be acquired in the absence of awareness or intention to learn, and thus has a hidden or tacit quality.

In this section, we begin by discussing the type of theoretical concept we consider tacit knowledge to be. Next, we describe the characteristic features of tacit knowledge and how it is distinguished from related concepts. Then, we consider how tacit knowledge is represented at different levels of abstraction. We present a cognitive model that relates the key features of tacit knowledge to the acquisition, storage, and retrieval of knowledge in and from memory.

3.1 Tacit knowledge as a theoretical concept

In research by Sternberg and his colleagues (Sternberg et al., 1993; Sternberg et al., 1995; Wagner and Sternberg, 1985), the term *tacit knowledge* has been used to characterize a type of knowledge, the possession of which distinguishes more from less practically-successful individuals. In order to understand better the theoretical concept of tacit knowledge, we begin with a distinction between nominal and natural concepts.

Nominal concepts are used attributively. For example, we use the term "bachelor" to attribute certain features (i.e., male, adult, unmarried) to some objects or persons. The instances of a nominal concept often share features that are both necessary (i.e., all valid instances must have these features) and sufficient (i.e., having these features is enough to qualify something as a valid instance). Membership in a nominal concept is "all or none"—either an instance possesses the critical features or it does not.

Natural concepts, in contrast, are used ostensively. For example, we use the term "furniture" to refer to objects that we view as equivalent (e.g., dresser, chair, table). The instances of a natural concept share characteristics features, but these features are not necessary or sufficient for membership. Membership in a natural concept is not "all or none," but rather instances are judged in terms of their strengths of resemblance to the concept. This means that some instances (those with high resemblance) will be judged as better examples of the concept than will other instances (those with low resemblance). For example, most people would agree that "arm chair" is a more typical example of the concept "furniture" than is "bean bag chair."

Tacit knowledge is a natural concept. It is used to denote a type of knowledge that is held together by the resemblance of items to one another and not by a set of individually-necessary and jointly-sufficient features. This lack of necessary and sufficient features does not mean that as a concept tacit knowledge is incoherent or meaningless. Two people may not be able to identify the critical features that all items of furniture share, but they can still agree that furniture exists and that a coffee table is furniture and a toaster oven is not.

Because tacit knowledge is a natural concept, we do not expect that judgments about what is and is not tacit knowledge will be "all or none." Rather judgments should depend on the item's strength of resemblance to the concept. Some knowledge will seem to represent a particularly clear example of tacit knowledge and other knowledge will seem marginal. For marginal items, individuals may disagree about whether the item is a valid instance of tacit knowledge. Given a high level of agreement among judges, the tacit quality of knowledge items can be determined with some degree of confidence.

We describe below three key features that are commonly shared by items of tacit knowledge. These features are used to judge the resemblance of items to the concept. In other words, items that possess these features are more likely to be characteristic of tacit knowledge.

3.2 The characteristic features of tacit knowledge

We identify three key features of tacit knowledge. These features of tacit knowledge relate to (a) the conditions under which it is acquired, (b) its cognitive structure, and

(c) the conditions of its use. First, tacit knowledge generally is acquired on one's own with little support from the environment (e.g., through personal experience rather than through instruction). Second, tacit knowledge is viewed as procedural in nature. It is associated with particular uses in particular situations or classes of situations. Third, because it generally is acquired through one's own experiences, tacit knowledge has practical value to the individual. We expand upon each of these features below.

3.2.1 Tacit knowledge typically is acquired without environmental support

Tacit knowledge generally is acquired on one's own. That is, it is acquired under conditions of minimal environmental support. By environmental support, we mean either people or media that help the individual to acquire the knowledge. As such, tacit knowledge tends to be unspoken, underemphasized, and poorly conveyed relative to its importance for practical success.

When people or media support the acquisition of knowledge, they facilitate three knowledge-acquisition components: selective encoding, selective combination, and selective comparison (Sternberg, 1988). When an individual is helped to distinguish more from less important information (selective encoding), to combine elements of information in useful ways (selective combination), and to identify knowledge in memory that is relevant to the present situation (selective comparison), the individual has been supported in acquiring knowledge. In performing real-world tasks, individuals often must engage in these processes on their own in order to make sense of and respond to situations. The resulting knowledge may reflect the use of these processes, but the individual may not be able to express how the knowledge was acquired.

3.2.2 Tacit knowledge is procedural

The second feature of tacit knowledge is its close association with action. Tacit knowledge takes the form of "knowing how" rather than "knowing that." Anderson (1983) has characterized these two respective types of knowledge as procedural and declarative. More precisely, procedural knowledge is knowledge that is represented in a way that commits it to a particular use or set of uses. It is knowledge that guides behavior, usually without being readily available to conscious introspection. People may not know they possess and/or may find it difficult to articulate such knowledge. We view procedural knowledge as a superset of tacit knowledge. All tacit knowledge is procedural, although not all procedural knowledge is tacit.

The characterization of tacit knowledge as procedural derives from our research. We have found that when individuals are queried about the knowledge they have acquired through their experiences, they often begin by articulating general rules in roughly declarative form (e.g., "a good leader needs to know people"). When these general statements are probed, the statements often reveal themselves to be more abstract or summary representations of a family of complexly specified procedural rules (e.g., rules about how to judge people accurately for a variety of purposes and under a variety of circumstances). These procedural rules, we believe, represent the characteristic structure of tacit knowledge and serves as the basis for identifying and measuring tacit knowledge.

We can represent tacit knowledge in the form of condition-action pairings:

IF <antecedent condition> THEN <consequent action>

For example, the knowledge of how to respond to a red traffic light could be represented as:

IF <light is red> THEN <stop>

Of course, the specification of the conditions and actions that make up proceduralized knowledge may be quite complex. In fact, much of the tacit knowledge that we have observed seems to take the form of complex, multicondition rules (production systems) for how to pursue particular goals in particular situations. In other words, tacit knowledge is more than a set of abstract procedural rules. It is context-specific knowledge about what to do in a given situation or class of situations. For example, knowledge about confronting one's superior might be represented in a form with a compound condition:

IF <you are in a public forum>

AND

IF <the boss says something or does something that you perceive is wrong or inappropriate >

AND

IF <the boss does not ask for questions or comments>

THEN <speak directly to the point of contention and do not make evaluative statements about your boss, staff or your peer's character or motives>

BECAUSE <this saves the boss from embarassment and preserves your relationship with him.>

3.2.3 Tacit knowledge is practically useful

The third characteristic feature of tacit knowledge is its instrumental value in attaining people's personal goals. The more highly valued the goal is, and the more directly the knowledge supports the attainment of the goal, the more useful is the knowledge. For example, knowing that seeking input from subordinates makes them feel valued is practically useful for those supervisors who want their subordinates to feel valued, but not practically useful for supervisors who do not value this goal.

We do not believe that practically useful knowledge must be acquired in any particular context or forum. Useful knowledge is, of course, acquired in classrooms, from experience on the job, through mentoring relationships, and through self-study. We distinguish practically useful knowledge not from formally acquired knowledge but, rather, from knowledge (however acquired) that is not relevant to the practical goals that an individual values.

3.2.4 Tacit knowledge involves coherent relations among its features

The three features of tacit knowledge, acquisition on one's own, procedural structure, and practical value, are related to one another in a non-arbitrary way. That is, we can explain why these features go together in the specification of a meaningful natural concept of tacit knowledge.

First, there is a natural correpondence between the features of procedural structure and practical value. Procedural knowledge tend to be practically useful—it contains within it the specification of how it is to be used. Declarative knowledge, in contrast, is not specific with respect to use and, as a consequence, may remain inert or unused. Therefore, procedural knowledge is more likely to be relevant in the pursuit of personally-valued goals.

Second, knowledge acquired under low environmental support is more likely to have practical value. When knowledge must be acquired on one's own, the probability increases that some individuals will fail to acquire it. When some individuals fail to acquire knowledge, those who succeed may gain a comparative advantage. This advantage is expected to be lower when the knowledge is highly supported by the environment (i.e., explicity and effectively taught) because more people would be expected to acquire and use it. At the same time, knowledge acquired through one's own experiences should have more personal relevance to the types of situations one encounters in everyday life.

Finally, we associate knowledge acquired through experience with knowledge that is procedural in structure. Because procedural knowledge is more difficult to articulate and more poorly conveyed relative to declarative knowledge, its acquisition is more likely to be a function of experiential learning. By the same token, knowledge acquired through experience is more likely to be related to action because originally it was obtained in the context of performing a practical, everyday task.

Each of these features is viewed as a continuous, rather than discrete, dimension of tacit knowledge. That is, knowledge is not categorized as either possessing or not possessing these features, but rather it is a matter of degree. Some knowledge may be more well-supported by the environment than other knowledge. Similarly, some knowledge may have more practical value to the individual than other knowledge. Knowledge that is closer to one end of the continuum is considered more representative of tacit knowledge.

3.3 What tacit knowledge is not

We have identified above the features that help describe what type of knowledge we consider tacit knowledge to be. It is helpful also to distinguish tacit knowledge conceptually from other related concepts such as job knowledge, general cognition, and performance.

3.3.1 Tacit knowledge is not synonymous with job knowledge

Schmidt and Hunter (1993) suggested that tacit knowledge is merely a type of job knowledge. Tacit knowledge and job knowledge are viewed more appropriately as overlapping concepts. First, some, but not all, tacit knowledge pertains to job-related activities. Tacit knowledge can pertain to any personally-valued activity, including academic and social activities; it is more than job knowledge. Second, some, but not all, job knowledge is tacit. Job knowledge includes declarative and procedural knowledge, with some of the latter characterized as tacit. Job knowledge may be explicit and readily verbalized, as in the rules for operating a lathe or the steps used to compute simple interest, or the knowledge may be tacit, as in knowing what package design will likely sell a product.

Measures of tacit knowledge have the potential to explain individual differences in performance that are not explained by traditional measures of job knowledge, which tend to assess more declarative, explicit forms of knowledge (see e.g., Schmidt and Hunter, 1998). Individual differences in the skill or inclination to acquire and use tacit knowledge make it a potentially useful construct for understanding intelligent behavior in real-world settings, as well as for predicting success in such settings.

3.3.2 Tacit knowledge is not a proxy for general cognition

The skill or propensity to acquire tacit knowledge is viewed as a dimension of practical cognition that conventional skill tests do not adequately measure. Overall cognitive ability tests and similar tests, which are intended to measure so-called general cognition (g), are composed of problems that can be characterized as largely academic or abstract.

As discussed earlier, academic problems are well-defined, abstract problems that do not necessarily reflect real-world tasks (Neisser, 1976; Sternberg, 1988, 1997a). Therefore, overall cognitive ability tests and similar tests measure problem-solving skills that are relatively different from the skills needed to solve everyday, practical problems. For this reason, we do not view measures of tacit knowledge as proxies for measures of academic cognition. Although general cognitive skill may support the acquisition and use of tacit knowledge in important ways, tacit knowledge is not reducible to academic cognition. Of course, it is an empirical question whether measures of tacit knowledge do in fact correlate with measures of crystallized cognition. This question is addressed in subsequent sections.

3.3.3 Tacit knowledge is not sufficient for effective performance

Although we do not consider tacit knowledge to be a proxy for general cognition, we do recognize that so-called *g* and other factors contribute to successful performance in many jobs, based on traditional criteria of success (such as performance ratings). The performance of many everyday tasks requires general academic cognition in (at least) the normative range, motivation to succeed, nontacit domain knowledge, and many other resources. We recognize and basically are in concurrence with the results of numerous meta-analyses that show the significant contribution of these variables to understanding performance (see Schmidt and Hunter, 1998). But we attempt to supplement these variables and improve upon conventional approaches to understanding, predicting, and improving performance in real-world settings.

Measures of practical cognition, like all measures of cognition, are, at best, indicators of the underlying cognitive functions we seek to understand. As such, we can talk about practical cognition, and more specifically tacit knowledge, at different levels of abstraction. That is, we can conceptualize tacit knowledge at the level of its cognitive representation, and at the level which it is measured in the behavior and articulated knowledge of the individual. We discuss these different levels of abstraction below.

3.4 Describing tacit knowledge at different levels of abstraction

Tacit knowledge can be conceptualized at qualitatively different levels of abstraction. At the lowest, least abstract level, tacit knowledge can be described as mentally-represented knowledge structures. We believe that these knowledge structures take the form of complex, condition-action mappings. At this level of description, tacit knowledge takes on its psychological reality and has its consequences for intelligent behavior.

Ideally, we would measure the possession of tacit knowledge directly at the level of its cognitive representation. However, we must infer possession of tacit knowledge from the knowledge that people articulate. When knowledge is articulated, often it is greatly simplified. That is, the complex knowledge structures that map sets of antecedent conditions onto consequent actions are summarized and abbreviated into general rules and procedures. It is at this level, that we measure people's tacit knowledge.

At a higher, more abstract level of description, tacit-knowledge items can be grouped into categories of functionally-related items. Describing tacit knowledge at this level adds value to the identification of tacit knowledge by highlighting the broad, functional areas or competencies that tacit knowledge represents. In other words, in addition to specific items of tacit knowledge, we can identify more generally the types of knowledge that are likely to be tacit.

3.4.1 Identifying and measuring tacit knowledge

Measuring tacit knowledge takes into account the realistic, contextualized quality of the knowledge. Responses to realistic problem situations are used as indicators of an individual's possession of tacit knowledge. Wagner and Sternberg (1985) devised a method of presenting scenarios to individuals that depict the types of problems they face in their given pursuits. These scenarios reflect the types of situations in which recognized domain experts have acquired knowledge characterized as "tacit." Because tacit knowledge is not readily articulated, we rely on observable indicators (e.g., responses to the scenarios) to assess whether an individual possesses knowledge characterized as tacit, and can apply that knowledge to the situation at hand. The responses reflect an individual's skill to recognize and take appropriate action in a given situation, and presumably, their procedural knowledge.

Deriving the information for these scenarios poses a challenge in that the tacit knowledge of domain experts must somehow be identified. Domain experts are appropriate sources for identifying tacit knowledge because in order to achieve their expert status, they likely have acquired knowledge that others have not (i.e., knowledge without direct support). As a subset of procedural knowledge that is not readily articulated, tacit knowledge is not likely to be elicited directly from individuals. However, since tacit knowledge is experience-based, we attempt to identify the knowledge in the recalled experiences of individuals. In other words, when individuals have difficulty expressing their action-oriented knowledge, we attempt to elicit memories for the particular episodes that produced that knowledge.

In the next section, we describe methods used to elicit examples of tacit knowledge from domain experts and to develop instruments to measure the acquisition and use of tacit knowledge within a given domain. The methods, which have been applied in domains ranging from education to military leadership, have evolved over the course of our tacit-knowledge research, resulting in a refined and detailed methodology for eliciting and measuring tacit knowledge. We devote the next section to describing this methodology as it plays an important role in understanding the findings from tacit-knowledge research and offers a tool for studying tacit knowledge in any domain.

4. Measuring tacit knowledge

One of the goals of our research is to show that tacit knowledge contributes to successful performance in a variety of domains. That is, we aim to establish a relationship between the possession of tacit knowledge and performance. But how does one proceed to develop a test to measure tacit knowledge? This section addresses the development of tools to measure the amount of tacit knowledge of various kinds that an individual has acquired. We begin by reviewing some approaches that have been used to measure the competencies considered to be relevant to the performance of real-world tasks, and contrast them with our knowledge-based approach. We then discuss what tacit-knowledge tests are intended to measure and offer a general framework for developing and validating such a test through the assessment of everyday situational judgments.

4.1 Methods of measuring real-world competencies

The tacit-knowledge approach to understanding practical cognition is based on several methods of measuring real-world competencies. These include the use of the critical-incident technique, simulations, and situational-judgement tests. We review briefly each of these methods and then discuss how the tacit-knowledge approach draws certain aspects from these methods.

4.1.1 Critical-incident technique

The critical-incident technique is an approach that seeks to identify the behaviors associated with effective performance (Flanagan, 1954). According to Flanagan, a critical incident describes the behavior, the setting in which the behavior occurred, and the consequences of the behavior. Critical incidents are generated by asking individuals, typically subject-matter experts, to provide examples of effective and ineffective behaviors. More specifically, individuals are asked, through interviews or open-ended survey questions, to describe several incidents that they, or someone else, handled particularly well, as well as several incidents that they, or someone else, handled poorly (Flanagan, 1954; McClelland, 1976). Boyatzis (1982) used a variation on the critical-incident technique, called the "behavioral event interview," in which he obtained behavioral incidents from individuals identified a priori as either high, medium, or low on effectiveness. He then examined the incidents generated from each group to identify traits and skills that distinguished between effective and ineffective managers.

The "critical incidents" generated from observations, interviews, or surveys are analyzed qualitatively to determine the nature of the competencies that appear important for success in a given task domain. The incidents typically are grouped on the basis of similar behavior content. For example, an incident that pertains to assigning a task to a subordinate and an incident about monitoring task completion by a subordinate might be grouped into a category of supervising subordinates. These categories are used to draw general conclusions about the behaviors that are characteristic of effective and ineffective performers.

Limitations of the critical-incident technique are that it assumes people can and will provide incidents that are critical to success in their particular jobs, and that qualitative analysis is sufficient for identifying the underlying competencies. However, the value of the critical-incident technique lies in identifying the strategies individuals use to perform various tasks, and in examining specific, situationally-relevant aspects of behavior. The critical-incident technique has been used successfully in the development of several performance assessment tools, including behaviorally anchored rating scales (BARS; e.g., Smith and Kendall, 1963) and situational-judgment tests (SJTs; e.g., Motowidlo, Dunnette, and Carter, 1990), the latter of which is described in more detail below.

4.1.2 Simulations

Simulations have been used as both assessment tools and as training methods. Simulations are aimed at assessing job behaviors directly. They involve observing people in situations that have been created to simulate aspects of the actual job situation. Responses to these simulations are considered to represent the actual responses that individuals would exhibit in real situations. Simulations can take the form of in-basket tests, situational interviews, group discussion, assessment centers, and situational-judgement tests. Motowidlo et al. (1990) distinguished between high-fidelity and low-fidelity simulations. In high-fidelity simulations, the stimuli presented to the respondent closely replicate the actual situation and the individual has an opportunity to respond as if they were in the actual situation. In low-fidelity simulations, the stimuli are presented in written or oral form and the individual is asked to describe how he or she would respond to the situation, rather than actually to carry out the behavior.

At the high-fidelity end of the continuum is the assessment center. Assessment centers present small groups of individuals with a variety of tasks, including in-basket tests, simulated interviews, and simulated group discussions (Bray, 1982; Thornton and Byham, 1982). The simulation approach has the advantage of more closely representing actual job performance. However, it is not always clear what aspects of the job should be chosen to simulate or how to evaluate performance.

In-basket tests have a moderate level of fidelity. In an in-basket test, the participant is presented with various materials (e.g., memos, financial reports, letters) and is asked to respond to them (Frederiksen, 1966; Frederiksen, Saunders, and Wand, 1957). The individual, however, has a limited amount of time to deal with the problems presented in the in-basket, giving him or her some of the constraints of actual job situations. Performance is evaluated based on how the items are handled. For example, does the participant respond to a letter from the Director of Finance requesting fourth-quarter financial records with complete and accurate information?

Situational-judgment tests have been considered low-fidelity simulations (see Motowidlo et al., 1990). Situational-judgment tests (SJTs) present descriptions of situations, typically work-related, in which a problem exists (see e.g., Chan and Schmitt, 1998; Legree, 1995; Motowidlo et al., 1990). The descriptions can be of actual situations or written to approximate actual situations in the domain of interest (e.g., a salesperson making a phone solicitation). Situations typically are selected on the bases of a critical-incident analysis. Following each situational description is a set of options (i.e., strategies) for solving the problem. Respondents are asked to indicate their endorsement of the options, either by selecting the best and possibly the worst from among a few strategies, or rating the effectiveness of each alternative. Traditionally, SJTs have been scored by awarding points based on the correct choice of the best and worst options (e.g., Motowidlo et al., 1990), or awarding points based on the percentage of experts who endorse the option (e.g., Chan and Schmitt, 1998).

4.1.3 Tacit-knowledge approach

The tacit-knowledge approach draws on aspects of the above approaches in order to measure the level of domain-specific, procedural knowledge that individuals have acquired from solving everyday problems. It is based on theoretical and empirical claims that the amount and organization of knowledge that experts possess differs from that of novices (see Chi et al., 1988; Ericsson, 1996; Ericsson and Smith, 1991), and that these knowledge differences reflect differences in the developed skills of experts and novices (Sternberg, 1998a; in press-a).

The tacit-knowledge approach relies on a critical-incident technique to identify examples of tacit knowledge acquired in solving real-word problems. That is, we interview domain experts to identify incidents that reflect important learning lessons, and ask them to express in their own words the knowledge gained from those situations. We do not rely solely on the individuals who provided the incidents to determine which items of knowledge are more or less effective. We use subsequent analyses to identify the items that are "critical" to performance.

The tacit-knowledge approach shares with the simulation approach the view that measuring practically relevant behavior in a test situation depends, in part, on the extent to which the task resembles those tasks found in everyday life. As such, we attempt to include sufficient detail in our measure to provide respondents with a realistic picture of the situation. However, we have relied primarily on a paper-and-pencil format to present this information rather than simulations for reasons of practicality, with the exception of our tacit-knowledge-acquisition task for sales (Sternberg et al., 1993). We have chosen to provide better coverage of the performance domain at the potential cost of lower fidelity. Future testing, however, is moving in the direction of more performance-based, high-fidelity assessment.

The tacit-knowledge approach is linked most closely to that of situational-judgment testing. We present situation descriptions, often based on actual situations of position incumbents, followed by several possible responses to those situations. The number of response options range between five and twenty. Individuals are asked to rate on a Likert scale the quality or appropriateness of each option for addressing the problem presented in the situation.

For example, in a hypothetical situation, an administrative assistant realizes that there is a factual error in a memo her boss has written and the memo needs to be sent out immediately. The boss is in a closed-door meeting. The respondent is asked to rate several options (usually on a 1 = low to 9 = high scale) for solving the problem. Examples of responses include (a) interrupting the meeting to show the boss the error, (b) fixing the error oneself and sending out the revision, and (c) fixing the error but waiting to send out the memo until the assistant can run it by the boss.

The set of ratings the individual generates for all the situations is used to assess the individual's tacit knowledge for that domain. Similar to SJTs, the scoring of tacit-knowledge tests often rely on the judgments of experts. In general, tacit-knowledge tests have been scored in one of three ways: (a) by correlating participants' responses with an index of group membership (i.e., expert, intermediate, novice), (b) by judging the degree to which participants' responses conform to professional "rules of thumb," or (c) by computing the difference between participants' responses and an expert prototype. To understand better what tacit-knowledge tests are designed to measure, we consider tacit knowledge as a measurement construct.

4.2 Tacit knowledge as a measurement construct

What are tacit-knowledge tests, and the items contained within them, intended to measure?

This question can be answered by considering a traditional distinction between achievement testing and cognition testing. In achievement testing, items are presumed to exemplify the measurement construct (e.g., knowledge of world history) but are not commonly viewed as predictors. For example, when an individual correctly answers a factual, multiple-choice question about world history, we assume that she possessed prior knowledge of either the fact in question or related facts that enabled her to rule out incorrect alternatives. We do not commonly view the history question as predictive of performance on other tests or tasks. In cognition testing, by contrast, items are presumed to predict performance but are not commonly viewed as exemplars of the measurement construct. For example, when an individual correctly solves a figural analogy problem, we do not assume that he possessed prior knowledge of the analogical relationship in question. However, we do view such analogy problems as predictive of performance on other tests and tasks of general mental skill.

Is a measure of tacit knowledge a cognition test or an achievement test? Having drawn a distinction between cognition and achievement testing, we must point out that neither type of test exists in a pure form (Sternberg, 1998a). All achievement tests measure underlying skills—if only the skills necessary to acquire and display mastery of the tested content—and so tend to have predictive value. Likewise, all cognition tests measure acculturated knowledge—if only the knowledge necessary to make sense of items and testing conventions—and so tell us something about the knowledge content of individuals rated high and low in general cognition. All of these tests measure a form of developing expertise (Sternberg, 1998a). Tacit-knowledge tests break down the (artificial) boundaries between achievement and skill testing.

Tacit-knowledge tests are everyday situational-judgment tests built on a theory of human cognition (Sternberg, 1995c). They are intended to measure both practical, experience-based knowledge and the underlying dispositions or skills that support the acquisition and use of that knowledge. Thus, scores on tacit-knowledge tests are expected to predict performance on tests or tasks that draw on either tacit knowledge or the mental skills that supported its development and use. These skills are hypothesized to differ from those implicated in the so-called "general factor" in human cognition commonly referred to as *g* and often approximately measured, in norm-referenced fashion, as IQ. Research by Sternberg and colleagues has produced support for the hypothesis that the skills associated with tacit-knowledge test performance are different than those associated with tests of *g* (Hedlund et al., 1999; Sternberg et al., 1993; Sternberg et al., 1995).

Because tacit-knowledge items are considered to measure both acquired knowledge and practical skill, we propose that tacit-knowledge tests have the potential to shed light upon (1) the content of tacit knowledge and (2) the events or experiences through which it was acquired. Few would contest that tacit-knowledge items reflect the knowledge of the respondents from whom the items were obtained (in the course of a "story-telling" exercise focusing on personal experiences). The items came from these respondents' memories and so must reflect the content of those memories. What remains to be determined is the degree to which tacit-knowledge items measure the acquisition and use of tacit knowledge by those who did not produce but, rather, endorsed or rated the items. This question is addressed by our numerous research studies in both civilian and military sectors, which we discuss in subsequent sections.

4.3 Developing everyday situational-judgment tests to measure tacit knowledge

We have developed tests to assess tacit knowledge for academic psychology, elementary-school teaching, business management, sales, entry-level jobs in organizations, college education, and military leadership. In this section we present a framework for developing tacit-knowledge tests of the format described above, a framework that is based on the techniques we have used to measure tacit knowledge in the various domains we have studied.

The development of tacit-knowledge inventories readily may be understood as a production process, beginning with the "raw materials" of experience-based tacit knowledge elicited from successful practitioners in a given domain and culminating in a revised and validated inventory. At each step in the development process, "value" is added through the conduct of research and analysis.

All of the phases are designed to support the development of assessment instruments based on (a) the theory and methods of tacit-knowledge research, and (b) the substantive knowledge in the domain of interest. Specifically, the steps are intended to aid in selecting the content that is most promising with respect to the goals of the assessment phase, that is, in measuring an individual's possession of tacit knowledge. The term *promising* is used here to refer to that subset of tacit knowledge with the highest probability of yielding or contributing to tacit-knowledge test questions that, taken together, constitute a valid measure of the underlying, domain-relevant tacit knowledge of respondents. This process was developed over the course of several research projects, and is applicable to the identification and assessment in tacit knowledge in any performance domain. We describe each stage in the process below, from the identification of exemplars of tacit knowledge to the construction of the final inventory.

4.3.1 Knowledge identification

We generally begin with a review of job-relevant literature (e.g., sales manuals, Army trade publications) to identify on a preliminary basis the experience-based, tacit knowledge for the relevant profession (e.g., salespersons, Army leaders). This review may suggest some of the content for use in a tacit-knowledge inventory, and may provide a preliminary taxonomy, or category framework, for organizing the knowledge. For example, in research with managers, Wagner and Sternberg (1986) proposed a framework of practically-intelligent behavior consisting of tacit knowledge about managing oneself, managing others, and managing one's career.

Typically, a review of the literature does not provide a sufficient number of examples of knowledge that meet our criteria or include enough detail from which to create tacit-knowledge questions of the format described above. We have found that the practical advice presented in the professional literature tends to be decontextualized and already converted to semantic knowledge. We also surmise that the politics of professional print may keep some truly tacit knowledge—knowledge that contradicts doctrine, for example—out of print altogether. Therefore, the next step is to conduct interviews with successful practitioners in the domain to generate a larger body of knowledge from which to draw in developing the tacit-knowledge inventories. We described here a method for conducting these interviews.

A method for eliciting tacit knowledge. In selecting individuals to interview, it is important to identify a sample that is likely to possess a certain amount of tacit knowledge. We seek to identify individuals who are both experienced and successful in their domain. Individuals who are more successful likely have acquired some important knowledge relevant to success that individuals who are less successful have not. Furthermore, individuals who are currently practicing in the domain of interest are more appropriate sources for understanding the tacit knowledge of that domain than are individuals who hold other positions (e.g., supervisor) or previously held the position of interest. The latter may consider different knowledge to be relevant, based on their different perspectives. Once a relevant pool of practitioners is identified, experts can be chosen either through nominations (e.g., by peers or superiors) or based on existing performance criteria (e.g., performance evaluation, salary). In research by Sternberg and his colleagues (Hedlund et al., 1999; Sternberg et al., 1993; Sternberg et al., 1995; Wagner, 1987), interviews were conducted with academic psychologists deemed successful based on their tenure and affiliation (e.g., full professors at Yale); business managers who were considered successful on the basis of their position in the company; salespersons who were successful in their sales performance; successful college students selected based on grades and school affiliation; and successful military leaders identified through a nomination process.

All of these experts were asked to consider what it takes to succeed in their respective domains and to provide typical performance-related situations and possible responses to those situations that exemplify tacit knowledge. In recent research, we have developed a structured interview in which participants are provided with more explicit instructions about the knowledge we seek to identify and which prompts them to elicit more in-depth responses.

We rely generally on a two-person interview team, with one person designated as the lead interviewer and the other the notetaker. The lead interviewer directs the interview and the notetaker takes written notes, asks for clarification, and, along with the lead interviewer, asks follow-up questions. The interviews also are taped when possible, with the interviewees' consent, so that questions and clarifications can be addressed once the interview is completed. It is also helpful for one of the interviewers to be familiar with the domain in order to understand any technical language or the jargon of the interviewee.

We present below a protocol for conducting the interviews. We use specific examples from our work with military leaders to illustrate the steps involved.

1. Introduction. When the participant arrives, members of the interview team introduce themselves and give a standardized introduction to the study and the interview. This introduction should state the purpose of the research, preempt likely misunderstandings, and orient the participant to the purpose of the interview. For example:

> *We are trying to understand the key lessons that leaders acquire from their experience on the job. If we can identify these lessons, we will try to find ways to use them to strengthen leader development efforts.*
>
> *This is not an evaluation of you as a leader. This is not a study comparing leaders from your organization to those from another organization.*
>
> *We want to identify specific examples of informal knowledge about leadership at your level. We want to find examples of things about leadership that are not written in books or taught in classes. Our belief is that this knowledge is often not discussed openly, but nevertheless is used by leaders as they meet the demands of their jobs. This knowledge may have been learned because of some challenge or problem you faced. It may have been acquired by watching someone else's successes or failures.*
>
> *We are not interested in the "party line" or "doctrine." We also are not interested in the purely technical things you learned from experience (e.g., how to tune up an engine). We are really interested in the problems and challenges you face and what you have learned about leadership at your level from these experiences.*

2. Request for stories. The purpose of the interview is to elicit stories or cases from the participants' experiences and to explore the unspoken, practical knowledge gained from or reflected in those cases. We ask participants, for example, to:

> *" Tell us a story about a leadership experience you have had as a leader in your current position from which you learned a lesson."*

The aim is to keep the focus on the stories rather than theories or generalizations about effective performance. In this way, the responses are more closely tied to the tacit-knowledge construct (i.e., in the knowledge based on personal, practical experience).

Because the values in the recalled experiences are sometimes unclear, we seek the participant's help in making sense of each story, and identifying the lesson associated with the story.

3. Follow-up questions. Follow-up questions are used to focus on key contextual variables in the stories (e.g., "Tell us more about the climate in your unit"); the goals and alternative courses of action reflected in the stories (e.g., "What exactly did you hope to accomplish?" and "What else did you consider doing at the time?"); and on identifying practical knowledge with broader applicability (i.e., "lessons learned") derived from the experiences described in the stories (e.g., "What do you think you learned from this experience?" and "How has this experience affected your approach to leadership?"). Once it appears that no more information can be gained from a story, the interviewer, given time allowances, may ask the participant to share another story from his or her experience.

At the completion of each interview, the notetaker summarizes the interview. An interview summary might contain the following information: (1) participant information (e.g., position, time in job, race, gender), (2) a summary of each story, (3) annotations to each story based on follow-up questions, and (4) any comments from the interviewer. It is useful for the notetaker and lead interviewer to review the summaries and resolve any disagreements over details or interpretations from the interview.

The identification of tacit knowledge does not end with the summarized interviews. Even with explicit instructions about what the interviewer is looking for, not all of the stories generated from the interviews provide examples of tacit knowledge. Therefore, the interview summaries are submitted to a panel of experts who are familiar with both the performance domain and the tacit-knowledge construct. These experts are asked to judge whether the interview summary represents knowledge that is intimately related to action, is relevant to the goals that the individual values, is acquired with minimal environmental support, and is relevant to performance in the domain under study (e.g., academic psychology, military leadership).

Products of the interviews. The products of the interviews are transcripts and summaries that contain numerous potential examples of tacit knowledge. These summaries serve two purposes in instrument development. First, tacit-knowledge "items" (essentially pieces of advice) may be extracted from the summaries and used in a number of later analyses. Second, the summaries themselves (consisting of stories that the professionals shared about their experiences) can be used directly in the construction of the inventory.

A useful interim step is to ask a panel of experts (e.g., members of the research team or practitioners familiar with the tacit-knowledge construct) to review the knowledge compiled from the interview summaries to ensure that it meets the criteria for tacitness. These criteria are that (1) the knowledge should have been acquired with little environmental support, (2) it should be related to action, and (3) it should have relevance to the goals that the person values. Often, upon further review, a knowledge example may be judged by experts to fail to meet one of these criteria. For example, consider the following story told by a military officer.

> *I had a lieutenant who was screwing up big-time. He would take sensitive items (e.g., weapons, night-vision devices, etc.) home. He even lost sensitive items. He lost a pistol, and rather than stop the mission and look for it, he continued on with the mission. As we all know, when you lose a sensitive item, you stop everything and look for it until you find it.*

The above story was deemed to lack the necessary criteria for tacitness. The interviewee indicated that the knowledge he referred to is generally known by leaders. It even may represent an official procedure. Therefore, we have no evidence that this knowledge is attributable to the officer's experience in dealing with sensitive items that are missing. On the other hand, consider a story from another officer about a similar issue.

> *It is important for a commander to know when to report bad news to the boss and when to withhold it. My unit had just completed a night move and had been in position for about two hours. A weapon was identified as missing around midnight. The section chief told me that the weapon was in the current position because he had seen it during the sensitive item checks. I talked to each member of the section and determined that the weapon was in the position. We looked for the weapon from about midnight until 0300 hours. During this time I chose not to notify the battalion commander because I was confident that the weapon would be found. However, a sensitive item report was due at 0400 hours, so, for ethical reasons, I notified the battalion commander at 0300 hours that the weapon was missing. I told the battalion commander what I had done so far and that I was confident that the weapon would be found at first light. He was not upset. We found the weapon within ten minutes after the sun came up and the battalion commander was pleased we followed the standard operating procedures for dealing with a missing weapon.*

In this story, the officer clearly expresses some knowledge he has acquired through previous experience in dealing with missing sensitive items (e.g., weapons). He has learned that, under some circumstances, it is best to hold off reporting a problem until it becomes necessary, so long as appropriate steps are taken to resolve the problem in the interim.

Coding the interview summaries. After determining which examples of knowledge meet the established criteria, it is useful to transform the summaries into a more usable form for the purpose of later analyses. We have used a format that is based on the procedural feature of our definition of tacit knowledge. That is, the knowledge is expressed as a mapping between a set of antecedent conditions and a set of consequent actions.

An item of knowledge is represented by one or more antecedent condition or "IF" statements, by one or more consequent action or "THEN" statements, and by a brief explanation or "BECAUSE" statement. The logical operators "AND" and "OR" are used in the coding to signal relationships of conjunction and disjunction, respectively. The operator "ELSE" is employed in the coding to connect sets of condition-action mappings into more complex procedures. Each individual piece of tacit knowledge is rewritten into this procedural form. This coding allows the researcher to analyze more readily the content of the tacit knowledge for the purpose of identifying categories of knowledge and selecting examples of knowledge that may be useful as items in a tacit-knowledge inventory. The result of this phase is a set of coded tacit-knowledge items.

The coded tacit-knowledge items then may be subjected to a sorting process to identify major categories of tacit knowledge. This sorting may entail asking a group of experts to organize the items according to categories of their own devising. The results of the independent sortings may be analyzed using hierarchical or other cluster analyses, a family of techniques for uncovering the natural groupings in a set of data (for more details regarding this technique, see Hartigan, 1975). This type of analysis may produce hierarchically organized clusters of items that can be expressed in the form of a tree. The clusters can be interpreted by experts and assigned labels that represent different categories of tacit knowledge. The categories may provide an indication of the major areas of learning that occur in one's respective field. The category framework is also useful in selecting items for test development that provide a broad representation of the performance domain.

4.3.2 Item selection

Although one may proceed to develop test questions directly from the tacit-knowledge items generated from the interviews, a further selection process may be necessary for a number of reasons. First, the interview study may yield too many items of tacit knowledge to include in a tacit-knowledge inventory of reasonable length, depending on the context in which the test might be used. Second, we cannot determine on the basis of the interviews alone what tacit knowledge is diagnostic of experience or predictive of effective performance in a given domain, or alternatively, what tacit knowledge is not related to these criteria. A manager, for example, may have learned that subordinates are more likely to come to her with problems if she leaves her door open. But the extent to which this practice contributes to her success is unclear. By leaving her door open she may become the repository for problems that are the responsibility of other managers, which may create a distraction for her from her job. Third, the results of the preliminary sorting of interview data may not be sufficient for determining the internal structure of the tacit-knowledge construct domain. That is, for the purposes of test construction, we would want further evidence of the structure of the performance domain to ensure the representativeness of our items. For the reasons above, we take an additional step to narrow down the pool of items from which test questions will be constructed.

The next step in the process of selecting items for instrument development is more quantitative than qualitative. It entails surveying job incumbents to assess the "quality" of each tacit-knowledge item. In order to develop a questionnaire that can be administered to job incumbents, the tacit-knowledge items may need to be condensed. For example, if we want professionals to evaluate 100 examples of tacit knowledge, it would be unreasonable to ask them to read 100 items in a very long and complex format. Therefore, it may become necessary to condense the items into briefer descriptions. Condensing the items involves extracting only the most important information and deleting unnecessary information. Attempts should be made to increase the comprehensibility of the items for the intended audience and to preserve the intent of the interviewee who provided the knowledge. The procedural structure that we consider to be characteristic of tacit knowledge is maintained in the rewriting of items.

The condensed items are compiled into a survey, which we refer to as a Tacit Knowledge Survey (TKS), which is in turn a situational-judgment test. A TKS differs from a tacit-knowledge inventory in that respondents are asked to rate the perceived quality of the tacit knowledge in the former, whereas they are asked to rate the quality of responses to the problem in the latter. Job incumbents can be asked to rate each item on a number of dimensions. We have used four seven-point scales that ask for the following judgments: (1) how *good* does the respondent think the advice is, (2) how commonly *known* does the respondent think the advice is, (3) how *often*, in the judgment of the respondent, do incumbents at the specified level face situations such as the one described, and (4) to what extent does the advice match the respondent's personal *concept* of job performance? Each of the scales is intended to provide a different sort of information about the tacit-knowledge item being rated. The "good" scale is intended to assess the overall quality of the knowledge being rated. The "known" scale is intended to assess one possible index of tacitness (i.e., on the theory that knowledge whose acquisition is not well supported by the environment may be less commonly known than other knowledge). The "often" scale is intended to assess the generalizability or applicability of knowledge items across job settings within the domain. Finally, the "concept" scale is intended to assess respondents' implicit theories of performance. Together, the four rating scales are intended to provide a comprehensive but non-redundant picture of each tacit-knowledge item for the purpose of evaluating each item's potential for development into tacit-knowledge test questions.

We are interested in items that are (1) rated as better advice by those considered to be successful in their domain; (2) not viewed as common knowledge by individuals in the domain; (3) representative of the situations faced by most individuals in the domain; and (4) a good fit to the concept of performance held by successful individuals in the domain. In order to identify items that are endorsed by individuals who are successful in a domain, we obtain data on a relevant performance criterion. In our research with military leaders, we obtained two criterion measures—experience and performance ratings. Experience was expressed in terms of expert-novice differences and performance was assessed using ratings of leadership effectiveness by other leaders. Responses to the TKS are analyzed along with the criterion measure to identify items that have promise for inclusion in the tacit-knowledge inventory. This analysis generates a number of item statistics that can be used in the selection process.

In our research, we used discriminant analysis to identify items that distinguish individuals with more from those with less experience (see Hedlund et al., 1999). In the discriminant analysis, a linear combination of the discriminating variables (e.g., TKS items) is derived that maximizes the divergence between groups (e.g., experienced/novice). The linear combination of the discriminating variables (the canonical discriminant function or CDF) can be tested for significance to determine if the set of variables distinguishes between groups. In addition, the correlations between discriminating variables and the CDF can be computed to assess the discriminating power of individual variables (e.g., TKS items).

We used point-biserial correlations between ratings on the items and ratings of effective performance to identify items that reflected the responses of effective performers. Item statistics such as these can be used, along with the category framework developed in the interview phase, to select items that have the most potential to explain successful performance and provide the best "coverage" of the tacit-knowledge domain.

4.3.3 Instrument construction

The "knowledge identification" and "item selection" phases generate several outputs that serve as materials for the final phase of "instrument construction." These outputs include: (a) interview transcripts and interview summaries, (b) the category framework derived from expert sortings and cluster analyses, (c) a set of item statistics for use in the selection of content for the inventories, and (d) the knowledge items retained on the basis of the category framework and item statistics from the questionnaire study. In the next phase of test development, preliminary inventory questions are constructed, using both selected knowledge items and the interview summaries from which they were drawn. A tacit-knowledge question consists of a situation description followed by several potential responses to that situation. Although the condensed tacit-knowledge item may serve to describe the situation, it is preferable to include the details from the original story to provide a richer, more in-depth problem description. Including more contextual and situation-specific information in the question provides the respondent with a clearer basis on which to evaluate the appropriateness of potential responses to the situation. The original story also provides a source for developing the response options to a question.

Once the researchers are satisfied with the form of the preliminary inventory, it is useful to circulate the inventory among experts in the domain. One method of obtaining feedback is to convene a focus group of experts to review and discuss the inventory. In our research, focus-group participants were given a brief introduction to the goals of the project and an explanation of the tacit-knowledge construct in non-technical language. They were asked to judge the construct-relatedness of the inventory questions by considering whether each question addresses knowledge gained through experience and fits the definition of tacit knowledge provided. In addition, focus group participants were asked to help "fill gaps" and "fix problems" in the inventory. In particular, they were asked to (a) provide additional, plausible response options for any question; (b) identify areas of confusion or lack of clarity; (c) identify problems of gender, racial, or ethnic bias; and (d) identify anything that did not "ring true" in the inventory questions.

The researcher can use the feedback from the focus group to revise the inventories. For example, inventory questions for which judgments of construct-relatedness are not unanimous (and positive) may be omitted from the inventory. Similarly, a response option or scenario feature that is objected to by two or more participants may be omitted. The focus group may suggest additional response options or scenario features, which can be added to the inventory. The final result of this test-development process is a revised tacit-knowledge inventory that can be administered to position incumbents and used to address further research questions, such as those regarding criterion-related construct validity.

4.3.4 Summary

The phases described above all are designed to support the construction of tacit-knowledge tests. The tacit-knowledge items acquired in the interview study form the raw materials for this construction process. During this process, the tacit-knowledge items are subjected to qualitative analysis (e.g., sorting into categories) and quantitative analysis (e.g., obtaining quality ratings). The various phases serve to address two basic questions about the pool of tacit-knowledge from which an instrument will be developed. First, which items are most promising for use in the construction of tacit-knowledge test questions? Second, what does the underlying structure represented by the tacit-knowledge items tell us about the structure of the construct domain so that we can design our tacit-knowledge tests to capture this domain? The result of this process is an inventory that has greater likelihood of possessing both internal and external validity. We discuss the issue of validity in the last part of this section.

4.4 Establishing the validity of tacit-knowledge inventories

An important part of developing any tests is to establish its construct validity. Unlike the development of many cognition-type tests, we do not rely solely on the qualifications that items should load heavily on a single factor and predict some external performance criteria as sufficient for concluding that a test measures the construct of interest. As Nunally (1970) and others have argued, such a "criterion-based" approach to test development is problematic and often produces measurement instruments of inferior quality. Specifically, such an approach may yield tests that suffer from low internal-consistency reliability, poor factor structure, and fragility with respect to criteria other than those on which the selection of items was based.

We rely on both theoretical and empirical justifications to establish the validity of tacit-knowledge tests. We use Messick's (1995) unified validity framework to show how tacit-knowledge theory and the phases of test development outlined above, contribute to the validity of our tacit-knowledge tests. Messick's framework treats the traditionally separate forms of validity (i.e., content, construct, and criterion) as aspects of a more comprehensive kind of construct validity. According to this framework, the essential goal of test validation is to support, through a combination of theoretical rationale and empirical evidence, the interpretation of test scores and the uses of scores under that interpretation.

4.4.1 The content aspect

The content aspect of validity refers to evidence that test content is relevant to and representative of the focal construct. It addresses the concerns that fall under the traditional heading of content validity. In the context of tacit-knowledge test development, the goal of *construct relevance* calls for tacit-knowledge test questions that are sensitive to knowledge of the type specified by the focal construct and insensitive to knowledge that falls outside the focal construct. A first step toward this goal is taken during the identification phase of test development, in interviews with job incumbents, when we orient participants toward personal experiences and away from formal principles or theory within their performance domains. A second step is taken in the item-selection phase when incumbents are asked to rate the quality of tacit-knowledge items. These ratings (i.e., item means and variances) may provide evidence regarding the relevance of tacit-knowledge items to the underlying construct. For example, tacit-knowledge items with low mean ratings (i.e., when respondents, on average, consider the knowledge represented in the item to be bad advice) may not be relevant to successful performance. And items with low variances (i.e., when respondents agree highly about the quality—good or bad—of the knowledge reflected in the item) may not reflect knowledge gained through personal experience if the knowledge is generally agreed upon as good. In addition to these steps, the goal of establishing construct relevance also is supported by asking domain experts, at various stages in the test development process, to judge the relevance of the items to the tacit-knowledge construct.

The goal of *construct representativeness* calls for tacit-knowledge items that are typical rather than atypical of knowledge-based items specified by the focal construct. An initial step toward this goal is taken in the identification phase by interviewing job incumbents that are representative of the range of specialty areas within the domain. For example, military leaders in the same position (e.g., platoon leader) may serve in one of many branches (e.g., infantry, engineering). Therefore, in our research, we sought to interview officers from these various branches to increase the representativeness of the knowledge that was elicited. A second step is taken during the item-selection phase, when participants are asked to rate how "often" a situation presented in a tacit-knowledge item occurs. Items that receive both a low mean and small variance, for example, are ones that most incumbents agree occur almost never, and therefore may not be representative of the knowledge domain. The categories derived from cluster analyses of the tacit-knowledge items also provide a source for ensuring construct representativeness. Items can be chosen to represent each of the major categories of tacit knowledge, thus providing better coverage of the construct domain. Finally, at several points during test development, expert judgments are sought regarding the construct representativeness of the items. After an initial pool of potential tacit-knowledge items is obtained from the interviews, an expert panel is asked to judge the representativeness of each item. The experts are asked to eliminate items that are too narrow or technical in focus (e.g., how to safely store chemical weapons) and knowledge that is relevant to a small proportion of job incumbents (e.g., how to manage stress at work if you are a single mom). Experts again are asked to evaluate the representativeness of the items after preliminary test questions have been developed.

4.4.2 The substantive aspect

The substantive aspect of validity refers to the theoretical rationale behind tacit knowledge and its relationship to task (test) performance. A step toward the goal of substantive validity is provided by our cognitive model and the characterization of tacit knowledge presented in the sixth section. The model illustrates how tacit, procedural knowledge is acquired and how it comes to be applied in solving everyday problems. The model also helps to illustrate how tacit knowledge confers a performance advantage (relative to that conferred by nontacit, procedural knowledge) in people's skill to respond to contextualized problems of realistic complexity. The characteristic features of tacit knowledge (i.e., acquisition on one's own, procedural nature, and instrumental value) further highlight its potential contribution to successful performance. The cognitive model of tacit knowledge, on which the identification and measurement of tacit knowledge is based, provides a theoretical rationale for tacit-knowledge test performance and, as such, directly serves the goal of substantive validity. Substantive validity also may be supported by showing, through empirical research, the extent to which participants draw on personally-experienced rather than received knowledge in performing everyday, real-world tasks.

4.4.3 The structural aspect

The structural aspect of validity refers to the level of fit between the internal structure of the test and the internal structure of the construct domain. It is related to the issue of construct representativeness we discussed earlier. A first step toward the goal of structural validity is taken by interviewing and eliciting knowledge from job incumbents in all areas that represent the performance domain. For example, in our study with military leaders, we interviewed officers in all three of the major branch categories within the Army (i.e., combat arms, combat support, combat service support). The goal of structural validity also is served by administering measurement instruments (e.g., the Tacit Knowledge Survey) to a wide variety of job incumbents. By using broad samples of job incumbents, we are able to avoid basing our analyses and test development on a restricted subset of the tacit-knowledge domain. Of course, the structural aspect of validity is addressed most directly through statistical techniques like cluster analysis and multidimensional scaling that identify the internal structure of the sample of items. By examining the internal structure we cast a wider net in our selection of tacit-knowledge items, and in so doing, we have improved our prospects for developing tacit-knowledge tests that mirror the structure of the construct domain (e.g., the domain of practical, action-oriented knowledge that individuals acquire from personal experience).

4.4.4 The generalizability aspect

The generalizability aspect of validity refers to the extent to which score properties and interpretations generalize across groups, settings, and tasks. The generalizability aspect includes concerns that traditionally fall under the heading of "reliability." In the context of tacit-knowledge test development, the goal of generalizability calls for tacit-knowledge test scores that generalize across (1) roles within the organization, (2) repeated administrations, and (3) alternate forms of the test. Test development efforts relevant to the content, substantive, and structural aspects of validity also are relevant to the generalizability aspect. In general, by seeking to specify and measure the construct, rather than merely pursuing correlation with an external criterion, we presumably increase the generalizability of score interpretations for our tacit-knowledge tests.

4.4.5 The external aspect

The external aspect of validity refers to the issue of criterion-related validity. That is, we seek to establish that the test relates to an external criterion. More specifically, the goal is to obtain evidence of convergent and discriminant validity. Establishing criterion-related validity entails showing that tacit-knowledge test scores correlate more highly (i.e., converge) with theoretically related constructs (e.g., performance) and correlate less highly (i.e., diverge) with theoretically distinct constructs (e.g., general cognition, formal job knowledge).

Test-development efforts to specify and measure the tacit-knowledge construct also support the goal of criterion validity. For example, job incumbents are asked to provide examples of important lessons they learned in the course of performing their job rather than knowledge they gained in school. These instructions increase the likelihood that the tacit-knowledge items obtained will be related to performance criteria and be distinct from formally-acquired knowledge. Research during the item-selection phase involves assessing more directly the relation of these items to external criteria. This step helps to identify tacit-knowledge items that are indicative of successful performance.

Beyond these efforts during test development, additional steps should be taken to provide evidence of convergent and discriminant validity. For tacit-knowledge tests, possible discriminant evidence would be that which discounts the effects of general cognition, reading comprehension, and formally-acquired knowledge on tacit-knowledge test scores. Evidence of convergent validity would include a correlation between tacit-knowledge test scores and variables such as perceived job effectiveness, degree and rate of career advancement, and performance on job-relevant tasks. To obtain such evidence requires conducting a validation study in which measures of these variables are administered to or obtained from individuals. For example, in our research with managers and military leaders, we administered the tacit-knowledge inventory along with a measure of general cognition and related constructs, and obtained various performance criteria, such as supervisor ratings, salary and productivity. Correlational and hierarchical regression analyses can be used to assess convergent and discriminant validity. Convergent validity is supported by a significant relationship between tacit-knowledge test scores and the performance criterion (e.g., supervisor ratings). Discriminant validity is supported by zero to moderate correlations with measures such as general cognition and general job knowledge, as well as the incremental validity of tacit-knowledge test scores beyond these measures.

4.4.6 The consequential aspect

The consequential aspect of validity refers to the value implications of the intended use of score interpretation as a basis for action. Because tacit-knowledge tests may be used for employee assessment and development, or even selection, it is important to consider how the knowledge included in those tests fits into the culture and rules of the organization. For example, if an item of tacit knowledge meets all the criteria discussed above (e.g., satisfies the definition of tacit, exhibits a strong positive correlation with effective performance), but it conflicts with the organizational culture (e.g., suggesting that females should be given less responsibility than males) or it involves disobeying a regulation (e.g., suggesting that financial figures should be fudged when information is unavailable), then it may be inappropriate to include the item in a tacit-knowledge test. Relying on experts to review the tacit-knowledge items throughout the test-development process helps to ensure that issues related to the consequential aspect of validity are addressed.

4.5 Summary

The goal of the test-development process outlined in this section is to support the construction of valid tacit-knowledge tests. Our theoretical model of tacit knowledge, described in the previous section, constitutes, we believe, a step in the direction of this goal. By elaborating on what we consider to be tacit knowledge at a theoretical level, we set the stage for a more detailed consideration of item content during the selection process and, in so doing, increase the substantive validity of our tests. The analysis of item ratings and performance data constitutes a second step towards measuring the construct. By identifying those items with the strongest association with performance criteria, we increase the probability that we will select items and construct test questions that embody the construct—given that tacit knowledge has clear benefits for performance. The analysis of the underlying structure by sorting items into categories constitutes a third step toward our goal. By examining the structure of the tacit-knowledge space (based on the data from our sample), we are able to make more informed decisions about the distribution of item content in our tacit-knowledge tests and, in so doing, increase the structural validity and generalizability of score interpretations. Finally, by conducting validation studies we provide support that tacit knowledge is relevant to understanding performance in the domain of interest and that it contributes to that understanding beyond traditional indicators of performance. In the next two sections we discuss the development and validation of tests to measure tacit-knowledge in civilian and military domains.

5. The role of practical cognition in everyday settings

Our program of research is based on the notion that there is more to successfully predicting performance than just measuring the so-called general factor from conventional psychometric tests of cognition (see Sternberg and Wagner, 1993). We propose that tacit knowledge, as an aspect of practical cognition, is a key ingredient of success in any domain. Of course, there are those who disagree with this position (see Jensen, 1993; Ree and Earles, 1993; Schmidt and Hunter, 1993, 1998), suggesting that individual differences in performance are explained primarily by general cognitive skill. Some proponents of using general cognitive skill tests argue further that the value of these tests are that they are applicable for all jobs, have lowest cost to develop and administer, and have the highest validity (e.g., Schmidt and Hunter, 1998). But even Schmidt and Hunter acknowledge that alternative measures such as work sample tests and job knowledge tests have comparable and perhaps even higher validities than general skill tests, and provide incremental prediction above the latter.

A program of research by Sternberg and his colleagues has conducted tacit-knowledge research with business managers, college professors, elementary-school students, sales people, college students, and general populations. This important aspect of practical cognition, in study after study, has been found generally to be uncorrelated with academic cognition as measured by conventional tests, in a variety of populations, occupations, and at a variety of age levels (Sternberg, et al., 1993; Sternberg et al., 1995; Wagner, 1987; Wagner and Sternberg, 1985). A major task of this tacit-knowledge research has been to identify the content of tacit knowledge and develop ways to measure the possession of tacit knowledge. Tacit-knowledge tests present a set of problem situations and ask respondents to rate the quality or appropriateness of a number of possible responses to those situations. (The format and development of tacit-knowledge tests were discussed in the previous section.) In this section, we review the tacit-knowledge studies that have been conducted in civilian settings and in the next section, we present a specific example of a tacit-knowledge project with military leaders.

5.1 Academic psychologists

One of the first studies in the program of tacit-knowledge research was conducted by Wagner and Sternberg (1985) with academic psychologists. Wagner and Sternberg developed a test of tacit knowledge for academic psychologists based on interviews with five full professors and administered the test to three groups. The first group consisted of 54 faculty members from 20 psychology departments, identified as either among the top fifteen nationally ranked colleges or outside the top fifteen. The second group consisted of 104 psychology graduate students from the same departments as the faculty members. The third group consisted of 29 Yale undergraduates. Each participant was given 12 work-related situations, each with from 6 to 20 response options. For example, one questions described a second-year assistant professor who in the past year had published two unrelated empirical articles, who had one graduate student working with him, and who had not yet received external funding. His goal was to become a top person in his field and get tenure in his department. Participants were asked to rate on a scale from 1 to 9 the value of several pieces of advice regarding what the professor could do in the next two months. Examples of advice include: (1) improve the quality of his teaching, (2) write a grant proposal, (3) begin a long-term research project that might lead to a major theoretical article, (4) concentrate on recruiting more students, (5) serve on a committee studying university-community relations, and (6) begin several related short-term projects, each of which may lead to an empirical article.

Responses to the test were scored by correlating ratings on each item with an index variable for group membership (1 = undergraduate, 2 = graduate student, 3 = faculty member). A positive correlation between item and group membership indicated that higher ratings on the item were associated with more expertise, and a negative correlation indicated the opposite. Wagner and Sternberg (1985) validated the test using several criteria. They obtained from faculty members citation rates, the number of publications, number of conferences attended in the last year, number of conference papers presented, distribution of time across teaching and research, academic rank, year Ph.D. was obtained, and level of institutional affiliation (high or low). For undergraduates, they obtained scores on the Verbal Reasoning section of the Differential Aptitude Test (Bennett, Seashore, and Wesman, 1974).

Wagner and Sternberg (1985) found that tacit-knowledge test scores correlated significantly, and positively, with number of publications (.33), number of conferences attended (.34), rated level of institution (.40), and proportion of time spent in research (.39). For the undergraduates, tacit-knowledge tests scores did not correlate significantly with verbal-reasoning scores ($r = -.04$, ns).

In a follow-up study by Wagner (1987), a revised version of the test was administered to 91 faculty, 61 graduate students, and 60 Yale undergraduates. The revised test contained 12 situations with 9 to 10 response options. Wagner obtained ratings for both conceptions of what the person would do in their actual job and what they would do in their ideal job. Scores were obtained for the overall test, and for six subscales that crossed three kinds of tacit knowledge: tacit knowledge about managing oneself, managing others, and managing tasks, with two orientations of tacit knowledge: local (pertaining to the situation at hand) versus global (pertaining to a bigger picture) tacit knowledge.

A different scoring method was used than in Wagner and Sternberg (1985). An expert profile was created by administering the test to a sample of professors who were nominated as high on practical cognition. A distance score (d^2) was computed between the participant's ratings and the mean of the experts' ratings. The mean d^2 values for the three groups were 339 for faculty, 412 for graduate students, and 429 for undergraduates, indicating that tacit knowledge increased, on average, with level of experience (a smaller value representing greater tacit knowledge). There were exceptions in each group, however, suggesting that what mattered was not merely experience by what one has learned from experience.

Wagner then examined the relationship of tacit knowledge with the same criterion measures that were used in Wagner and Sternberg (1985). Because the tacit-knowledge test was scored using a distance measure, a lower distance, or smaller value, represents better tacit-knowledge score. Therefore, negative correlations reflect a positive association between tacit-knowledge scores and the criterion.

For the actual-job ratings, significant correlations were obtained between tacit-knowledge scores and ratings of department (-.48), number of citations (-.44), number of publications (-.28), proportion of time spent on research (-.41), and number of papers presented. The correlations for ideal-job ratings were slightly lower, but comparable. Again, the tacit-knowledge scores did not correlate with verbal-reasoning skill. Wagner did find significant intercorrelations among the six subscales, ranging from .2 to .4. He interpreted these correlations to indicate a weak general factor for tacit knowledge, a factor that appears to be distinct from the general factor measured by traditional cognition tests.

5.2 Business managers

Wagner and Sternberg (1985) and Wagner (1987) conducted studies with business managers in parallel to the studies with academic psychologists described above. That is, they involved similar methods but with a different performance domain.

Wagner and Sternberg (1985) developed a tacit-knowledge test for business managers based on interviews with 5 experienced and successful mid-level managers. The test consisted of 12 work-related situations with 9 to 20 response options and was administered to 54 managers (19 of whom were from among top 20 Fortune 500 companies), 51 graduate students from 5 business schools varying in prestige, and 22 Yale undergraduates. The criteria obtained for the managers included status in or outside the top Fortune 500 companies, number of years of management experience, number of years of formal schooling, salary, number of employees supervised, and level of job title. Undergraduates completed the DAT Verbal Reasoning subtest.

Responses to the test were scored by correlating ratings on each item with an index variable for group membership (1=undergraduates, 2=business school graduate students, 3=business managers). Wagner and Sternberg found significant correlations between tacit-knowledge and company level (.34), number of years of schooling (.41), and salary (.46). For the undergraduates, the correlation between tacit-knowledge scores and verbal-reasoning skill was not significant (.16), and again indicated that the tacit knowledge test was not a proxy for a traditional general cognition test.

In the second study, Wagner (1987) administered the test to 64 business managers, 25 business graduate students, and 60 Yale undergraduates. The distance scoring method, described above, was used. An expert profile was created from the responses of 13 business executives from Fortune 500 firms. The mean tacit-knowledge scores were 244 for business managers, 340 for business graduate students, and 417 for undergraduates, indicating greater tacit knowledge with experience. Correlations with the criterion measures were lower than those for academic psychologists. However, a significant correlation was obtained between tacit-knowledge scores and the number of years of management experience (-.30). Other correlations were in the predicted direction, but not significant. There was no significant correlation between tacit-knowledge scores and verbal-reasoning scores. And again, the six subscales generally correlated significantly with one another, with values ranging from .2 to .5, indicating a weak general factor for tacit knowledge.

In this study, the undergraduate participants completed the tacit-knowledge tests for both academic psychologists and business managers. The correlation between scores on the two tests was .58 and highly significant. Wagner concluded that not only do the subscales of the tacit-knowledge test correlate within a domain, but tacit-knowledge also appear to correlate across domains.

5.3 Center for creative leadership study

Further research on what later was formalized as the Tacit Knowledge Inventory for Managers (TKIM; Wagner and Sternberg, 1991) was conducted with a sample of 45 business executives who were participants in a Leadership Development Program (LPD) at the Center for Creative Leadership (Wagner and Sternberg, 1990). The purpose of the study was to validate the test against a managerial simulation and assess its discriminant validity with a variety of psychological measures. Wagner and Sternberg (1990) administered the TKIM with 9 work-related scenarios, each with 10 response options. Participants also completed, as part of the program, the Shipley Institute for Living Scale, a cognition test; the California Psychological Inventory, a self-report personality inventory; the Fundamental Interpersonal Relations Orientation-Behavior (FIRO-B), a measure of desired ways of relating to others; the Hidden Figures Test, a measure of field independence; the Myers-Briggs Type Indicator, a measure of cognitive style; the Kirton Adaptation Innovation Inventory, a measure of preference for innovation; and the Managerial Job Satisfaction Questionnaire. The participants' behavior was also assessed on two managerial simulations.

Beginning with zero-order correlations, the best predictors of managerial performance on the simulation were tacit knowledge (r = -.61, p < .001) and overall cognitive ability (r = .38, p < .001). (The negative correlation for tacit knowledge reflects the deviation scoring system used, in which better performance corresponds to less deviation from the expert prototype and thus to lower scores.) The correlation between tacit knowledge and overall cognitive ability was not significant (r = -.14, p > .05).

Hierarchical regression analyses were performed to examine the unique predictive value of tacit knowledge when used in conjunction with the various other cognition and personality tests. For each hierarchical regression analysis, the unique prediction of the TKIM was represented by the change in R^2 from a restricted model to a full model. In each case, the restricted model contained a subset of all the measures, and the full model was created by adding the TKIM to the equation. A significant change in R^2 indicated that the predictive relation between tacit knowledge and the simulation performance was not accounted for by the set of predictors in the restricted model.

In every case, tacit knowledge accounted for a significant increases in variance. In addition, when tacit knowledge, IQ, and selected subtests from the personality inventories were combined as predictors, nearly all of the reliable variance in the criterion was accounted for. These results support the strategy of enhancing validity and utility by supplementing existing selection procedures with additional ones. They also suggest that the construct of tacit knowledge cannot readily be subsumed by the existing constructs of cognitive skill and personality represented by the other measures used in the study.

5.4 Salespeople

Two studies were conducted by Wagner, Rashotte and Sternberg (1994; see also Wagner, Sujan, Sujan, Rashotte, and Sterberg, 1999) with salespeople. The objective of the first study was to develop and validate a "rules-of-thumb" approach to measuring tacit knowledge. Previous studies relied on empirical scoring, using either the correlation between items and an index of group membership or the deviation from an expert profile. Wagner et al. sought to identify a more objective, expert-based scoring method based on the rules of thumb that salespeople use to optimize their performance.

Based on a interviews, literature on sales, and personal experience, these investigators generated a list of rules of thumb for salespeople. The rules of thumb were divided into several categories, such as setting sales goals, handling the customer who stalls, attracting new accounts, and handling the competition. In the category of attracting new accounts, examples of rules of thumb included (1) be selective in regard to whom you direct your promotion efforts and (2) ask your customers to provide leads to new accounts.

The sample consisted of two groups. The first group consisted of salespeople with an average 14 years sales experience. The second group consisted of 50 undergraduates at Florida State University. The participants were administered eight sales scenarios, with 8 to 12 response options constructed by the rules-of-thumb approach. The options included accurate representations of the rules of thumb as well as weakened or distorted versions of them. Responses were evaluated based on the extent to which participants preferred the actual or distorted versions of the rules of thumb. In addition to the sales test, the undergraduates completed the DAT Verbal Reasoning test.

Participants were asked to rate the appropriateness of each strategy for addressing the problem. Points were awarded based on the participant's endorsement of the actual rules of thumb. Wagner et al. found that scores on the tacit-knowledge test improved with experience. The average score for salespeople was 209 versus 166 for undergraduates. The total scores for undergraduates were uncorrelated with verbal-reasoning test scores.

In the second study, measures of sales performance were obtained in addition to tacit-knowledge test scores. Participants included 48 life-insurance salespeople with an average of 11 years sales experience and 50 undergraduates at Florida State University with no sales experience. Participants in both groups completed the TKIS, and undergraduates completed the DAT Verbal Reasoning test. In addition, the investigators obtained from the salespeople data on the number of years with the company, number of years in sales, number of yearly quality awards, yearly sales volumes and premiums, college background, and business education.

Tacit knowledge again increased with experience, with the scores 165 and 206 for undergraduates and salespeople respectively. Significant correlations were obtained between tacit-knowledge scores and number of years with the company (.37), number of years in sales (.31), number of yearly quality awards (.35), and business education (.41). When local and global scores were also computed, Wagner et al. found that global tacit-knowledge scores also correlated significantly with yearly sales volumes and premiums (*r*s ranging from .26 to .37). The tacit-knowledge scores again did not correlate significantly with verbal-reasoning scores.

5.5 Air force recruits

In a study carried out at the Human Resources Laboratory at Brooks Air Force Base under the supervision of Malcolm Ree, Eddy (1988) examined relations between the TKIM and the Armed Services Vocational Aptitude Battery (ASVAB) for a sample of 631 Air Force Recruits, 29 percent of whom were females, and 19 percent of whom were members of a minority group. The ASVAB is a multiple-aptitude battery used for selection of candidates into all branches of the United States Armed Forces. Prior studies of the ASVAB suggest that it is a typical measure of cognitive skill, with correlations between ASVAB scores and other cognitive skill measures of about .7. Factor-analytic studies of the ASVAB also suggest that it appears to measure the same verbal, quantitative, and mechanical skills as the Differential Aptitude Tests, and the same verbal and mathematical knowledge as the California Achievement Tests.

Eddy's (1988) study showed small correlations between tacit knowledge and ASVAB subtests. The median correlation was -.07, with a range from .06 to -.15. Of the 10 correlations, only two were significantly different from zero, despite the large sample size of 631 recruits. A factor analysis of all the test data, followed by oblique rotations, yielded the usual four ASVAB factors (vocational-technical information, clerical/speed, verbal skill, and mathematics) and a distinct tacit-knowledge factor. The factor loading for the TKIM score on the tacit-knowledge factor was .99, with a maximum loading for scores on the four ASVAB factors of only .06. Upon oblique rotation, the four ASVAB factors were moderately intercorrelated, but the correlations between the tacit knowledge factor and the four ASVAB factors were near zero (.075, .003, .096, .082).

An additional point about these results concerns the possibility that measures of tacit knowledge might identify potential managers from nontraditional and minority backgrounds whose practical knowledge suggests that they would be effective managers, even though their performance on traditional selection measures such as cognition tests does not. Eddy (1988) did not report scores separately by race and sex, but did report correlations between scores and dummy variables indicating race and sex. Significant correlations in the .2 to .4 range between ASVAB subtest scores and both race and sex indicate that on the ASVAB, minority-group members scored more poorly than majority group members, and women scored more poorly than men. Nonsignificant correlations between tacit knowledge and both race (.03) and sex (.02), however, indicate comparable levels of performance on the tacit-knowledge measures between minority and majority-group members and between females and males.

5.6 Managers across organizational levels

In a study focusing on the development of tacit knowledge over the managerial career, Williams and Sternberg (cited in Sternberg et al., 1995) constructed a measure of both a general and a level-specific tacit-knowledge. They obtained nominations from superiors for “outstanding” and “underperforming” managers at the lower, middle, and upper levels in four high-technology manufacturing companies. This approach allowed them to delineate the specific content of tacit knowledge for each level of management (lower, middle, and upper) by examining what experts at each level knew that their poorly-performing colleagues did not.

Williams and Sternberg identified specialized tacit knowledge for each of the three management levels and found that this knowledge was differentially related to success. These results were derived from comparing responses of outstanding and underperforming managers within each management level on level-specific tacit-knowledge inventories. Within the domain of intrapersonal tacit knowledge, knowledge about how to seek out, create, and enjoy challenges is substantially more important to upper-level executives than to middle- or lower-level executives. Knowledge about maintaining appropriate levels of control becomes progressively more significant at higher levels of management. Knowledge about self-motivation, self-direction, self-awareness, and personal organization is roughly comparable in importance at the lower and middle levels, and somewhat more important at the upper level. Finally, knowledge about completing tasks and working effectively within the business environment is substantially more important for upper-level managers than for middle-level managers, and substantially more important for middle-level managers than for lower-level managers. Within the domain of interpersonal tacit knowledge, knowledge about influencing and controlling others is essential for all managers, but especially for those at the upper level. Knowledge about supporting, cooperating with, and understanding others is extremely important for upper-level executives, very important for middle-level executives, and somewhat important for lower-level executives.

In addition, Williams and Sternberg examined the relationship of tacit knowledge with several criteria across levels. They found that tacit knowledge was related to the following measures of managerial success: compensation (r = .39, p < .001), age-controlled compensation (r = .38, p < .001), and level of position (r = .36, p < .001). These correlations were computed after controlling for background and educational experience. Tacit knowledge was also moderately associated with enhanced job satisfaction (r = .23, p < .05).

These investigators further found that age, years of management experience, and years in current position were unrelated to tacit knowledge. The lack of a correlation of tacit knowledge with years of management experience suggests that it is not simply experience that matters, but perhaps what a manager learns from experience. A manager's years with current company was negatively related to tacit knowledge ($r = -.29$, $p < .01$), perhaps indicating that ineffective managers stayed around longer than effective managers. The number of companies that a manager had worked for was positively correlated with tacit-knowledge scores ($r = .35$, $p < .001$). Years of higher education was highly related to tacit knowledge ($r = .37$, $p < .001$), as was self-reported school performance ($r = .26$, $p < .01$). Similarly, college quality was related to tacit knowledge ($r = .34$, $p < .01$). These results in conjunction with the independence of tacit knowledge and overall cognitive ability suggest that tacit knowledge overlaps with the portion of these measures that are not predicted by overall cognitive ability.

Williams and Sternberg also performed hierarchical regression analyses to examine whether tacit knowledge contained independent information related to success that was distinct from that provided by background and experience. The pattern of results was similar across analyses. In the regression analysis predicting maximum compensation, the first variable entered in the regression equation was years of education, accounting for 19% of the variance ($p < .001$). The second variable entered was years of management experience, accounting for an additional 13% of the variance ($p < .001$). The third and final variable entered was tacit knowledge, accounting for an additional 4% of the variance ($p = .04$), and raised the total explained variance to 36%. In the regression predicting maximum compensation controlling for age, the number of years of education was entered into the equation first, accounting for 27% of the variance ($p < .001$). And second, tacit knowledge was entered, explaining an additional 5% of the variance ($p = .03$). This final regression demonstrates the value of tacit knowledge to managers who are relatively successful for their age.

5.7 College students

Williams and Sternberg (cited in Sternberg et al., 1993) studied the tacit knowledge of college students. They asked 50 Yale undergraduates the question: "What does it take to succeed at Yale that you don't learn from textbooks?" and used the responses to develop a tacit-knowledge inventory for college students. The inventory consisted of 14 situations and asked respondents to rate the quality of several options on a 1 to 9 scale. For example, one question described a student enrolled in a large introductory lecture course. The class requirements included three exams and a final. Participants were asked to rate how characteristic of their behavior it was to spend time doing various activities, such as (1) attending class regularly, (2) attending optional weekly review sections with a teaching fellow, (3) reading assigned text chapters thoroughly, (4) taking comprehensive class notes, and (5) speaking with the professor after class and during office hours.

The criteria were two indices: an academic index and an adjustment index. The academic index was a composite of high school GPA, college GPA, SAT scores, and CEEB achievement test scores. The adjustment index was a composite of a measure of happiness in college, a measure of self-perceived success in college, a measure of self-perceived success in using tacit knowledge, a measure of the extent of benefit each participant had experienced from acquiring tacit knowledge, and a measure of the rated closeness of the college to the participant's ideal college.

The academic and adjustment indices were not significantly correlated (-.09). Individual items of tacit knowledge correlated differently with the academic index and the adjustment index. The academic index was correlated with the perceived importance of maintaining a high GPA (.42); doing extra reading and school work not specifically assigned (.27); not attending optional weekly review sections (.23); not skimming required reading the morning before class (.37); not preparing a brief outline of points to raise in class discussion (.31); not helping friends with their assignments (.34); not behaving consistently from situation to situation (.25); finding it uncharacteristic to accept pressure and stress as parts of life (.30); finding it uncharacteristic to stand up for oneself (.34); and finding it uncharacteristic to play a sport or exercise regularly (.45).

Items that correlated significantly with the adjustment index included beliefs that professors value a clear, direct writing style, good organization of thoughts and ideas, and creative or unusual ideas (.38); beliefs that professors value papers that bring in outside interests or material (.27); beliefs that it is important sometimes to take on too many responsibilities at once (.31); seeking advice from several faculty in addition to one's own professors (.31); taking classes that permit occasional absences (.36); being positive and looking on the bright side of life (.42); not being intimidated (.33); being flexible (.27); maintaining a strong sense of confidence and independence (.37); not worrying unnecessarily or destructively (.31); knowing how to make oneself happy (.32); and not letting small disappointments affect one's long-term goals (.29).

Williams and Sternberg also obtained prediction of academic and adjustment indices with subsets of items from the tacit-knowledge inventory. Four items (not preparing an outline of points to raise in class discussion; maintaining a high GPA; not helping friends with assignments; and not playing a varsity or intramural sport) were predictive of the academic index, with an overall R^2 of .43. Six items (believing professors value a clear, direct writing style; maintaining a strong sense of confidence and independence; standing up for oneself; sometimes taking on too many responsibilities at once; seeking advice from faculty in addition to the course instructor; and taking classes that permit occasional absences) were predictive of the adjustment index, with an overall R^2 of .63. This study showed that tacit knowledge is important not only in occupational settings, but in school settings as well.

5.8 Conclusions from the tacit-knowledge research program

We organize a discussion of the findings from the tacit-knowledge research around four main issues: (a) the relationship of tacit knowledge to experience; (b) the relationship of tacit knowledge to general cognition; (c) tacit knowledge as a general construct; and (d) the relationship of tacit knowledge to performance.

5.8.1 Tacit knowledge and experience

In most of the studies reviewed above, tacit knowledge was found to relate to experience, indicated either by group membership (expert versus novice), or the number of years in one's current position.

In several studies, Sternberg and his colleagues showed that individuals with less experience in a given domain exhibit lower tacit-knowledge scores (Wagner, 1987; Wagner and Sternberg, 1985; Sternberg et al., 1993). In Wagner and Sternberg (1985), for example, group differences were obtained between business managers, business graduate students, and undergraduates on 39 of the response-item ratings on a tacit-knowledge test for managers, with a binomial test of the probability of finding this many significant differences by chance yielding $p < .001$. Comparable results were obtained with Yale undergraduates, psychology graduate students, and psychology faculty on a tacit-knowledge test for academic psychologists. In addition, Wagner (1987) found that business managers obtained the highest tacit knowledge scores followed by business graduate students, and undergraduates, with comparable results obtained in a study with psychology professors, psychology graduate students, and undergraduates. Wagner et al. (1994) also found that scores on a tacit-knowledge test for salespeople correlated significantly with number of years of sales experience.

Williams and Sternberg (cited in Sternberg et al., 1995), however, did not find significant correlations between several experience-based measures, including age, years of management experience, and years in current position, and tacit-knowledge scores. But they did find that the importance of specific pieces of tacit knowledge varied across organizational level. Their findings suggest that it may not simply be the amount of experience but what a manager learns from experience that matters to success.

5.8.2 Tacit knowledge and general cognition

In proposing a new approach to measuring cognition, it is important to show that one has not accidentally reinvented the concept of "*g*," or so-called general skill, as measured by traditional cognition tests. We do not dispute the relevance of general cognitive skill to performance. Schmidt and Hunter (1998) have shown that *g* predicts performance in a number of domains. Our aim is to show that tacit-knowledge tests measure something in addition to *g*. In all the above studies in which participants were given a traditional measure of cognitive skill, tacit-knowledge test scores correlated insignificantly with *g*.

The most consistently used measure of *g* in the above studies was the Verbal Reasoning subtest of the DAT. The absolute values of the correlations between tacit knowledge and verbal reasoning ranged from .04 and .16 with undergraduate samples (Wagner, 1987; Wagner and Sternberg, 1985) and .14 with a sample of business executives (Wagner and Sternberg, 1990).

One potential limitation of these findings is that they were obtained with restricted samples (e.g., Yale undergraduates, business managers). However, similar support for the relationship between tacit knowledge and *g* was found in a more general sample of Air Force recruits studied by Eddy (1988). The correlations between scores on the TKIM and ASVAB scales were modest, and none of the four ASVAB factors correlated significantly with the tacit-knowledge factor.

Tacit-knowledge tests may also be a better predictor than measures of personality, cognitive style, and interpersonal orientation as suggested by the findings from the Center for Creative Leadership study (Wagner and Sternberg, 1990). Sternberg and Grigorenko recently developed a test of common sense for the work place (e.g., how to handle oneself in a job interview) that predicts self-ratings of common sense but not self-ratings of various kinds of academic skills. The test also predicts supervisory ratings at a correlational level of about 4.

Finally, there is evidence that tacit knowledge may even correlate negatively with measures of academic cognition and achievement in some environments. In a study in a rural village in Kenya, Sternberg et al. (in press) developed a test to measure children's tacit knowledge for herbal medicines used to treat various illnesses. Parasitic infections are endemic among this population, and knowledge of these medicines and how to use them is important to adaptation to the environment. This knowledge, however, is not acquired in the classroom, but rather in the community from family members and healers.

The tacit-knowledge test for herbal medicines consisted of brief stories describing the specific manifestations of a given illness and provided the child with options regarding how to treat the illness (see Sternberg et al., in press). The tacit-knowledge test, along with the Raven Colored Progressive Matrices test (Raven, 1958), the English Mill Hill Vocabulary Scale (Raven, Court, and Raven, 1992), Dholuo (home language) Vocabulary Scale, and school-based measures of English and math achievement, were administered to 85 children ages 12 to 15. The tests of academic cognition were all significantly and positively correlated with each other. Scores on the tacit-knowledge test correlated in a negative direction with all of the academic cognition tests, and showed a significant negative correlation with scores on the vocabulary tests. Tacit-knowledge scores also exhibited a significant negative correlation with English achievement. Sternberg et al. concluded that practical cognition, as manifested in tacit knowledge relevant to adaptation in daily life, may be distinct from the kind of academic cognition associated with school success. The negative correlation between tacit-knowledge scores and some of the academic-cognition measures supports the claim that expertise developed in one environment (e.g., school) may have limited application in other environments (e.g., home or community life). Thus, there is a growing body of evidence, in work, school and community settings, which suggests that tacit knowledge measures a distinct construct from general, academic cognition.

5.8.3 Tacit knowledge as a general construct

Although the kinds of informal procedural knowledge measured by tacit-knowledge tests do not correlate with traditional psychometric cognition, tacit-knowledge test scores do correlate across domains. Furthermore, the structure of tacit knowledge appears to be represented best by a single, general factor.

Wagner (1987) examined the structure of tacit knowledge inventory for managers. He performed two kinds of factor analyses on the tacit-knowledge scores of these business managers in his study. First, a principal-components analysis yielded a first principal component that accounted for 44 percent of the total variance, and 76 percent of total variance after the correlations among scores were disattenuated for unreliability. The 40 percent variance accounted for by the first principal component is typical of analyses carried out on traditional cognitive-skill subtests. Second, results of a confirmatory factor analysis suggested that a model consisting of a single general factor provided the best fit to the data. The results of both factor analyses suggested a general factor of tacit knowledge.

Similar analyses were performed on a measure of tacit knowledge for academic psychologists. Consistent with the manager study, the factor analytic results suggested a single factor of tacit knowledge within the domain of academic psychology. Wagner (1987) also examined the generalizability of tacit knowledge across domains by administering both tacit-knowledge measures (for business managers and academic psychologists) to undergraduates in his study. He obtained a significant correlation of .58 between the two scores, suggesting that in addition the existence of a general factor of tacit knowledge within a domain, individual differences in tacit knowledge generalize across domains. These findings lend support for a common factor underlying tacit knowledge; a factor that is considered to be an aspect of practical cognition

5.8.4 Tacit knowledge and performance

Finally, we have shown that tacitknowledge measures are predictive of performance in a number of domains, correlating between .2 to .5 with measures such as rated prestige of business or institution, salary, simulation performance, and number of publications.

These correlations, uncorrected for attenuation or restriction of range, compare favorably with those obtained for overall cognitive ability within the range of skills we have tested.

In studies with business managers, tacit-knowledge scores correlated in the range of .2 to .4 with criteria such as salary, years of management experience, and whether or not the manager worked for a company at the top of the Fortune 500 list (Wagner, 1987; Wagner and Sternberg, 1985). Wagner and Sternberg (1990) obtained a correlation of .61 between tacit knowledge and performance on a managerial simulation, and found that tacit-knowledge scores explained additional variance beyond overall cognitive ability and other personality and skill measures. In a study with bank branch managers Wagner and Sternberg (1985) obtained significant correlations between tacit-knowledge scores and average percentage of merit-based salary increase ($r = .48$, $p < .05$) and average performance rating for the category of generating new business for the bank ($r = .56$, $p < .05$).

Williams and Sternberg (cited in Sternberg et al., 1995) also found that tacit knowledge was related to several indicators of managerial success, including compensation, age-controlled compensation, level of position, and job satisfaction, with correlations ranging from .23 to .39.

Although much of the tacit-knowledge research has involved business mangers, there is evidence that tacit knowledge explains performance in other domains. In the field of academic psychology, correlations in the .4 to .5 range were found between tacit-knowledge scores and criterion measures such as citation rate, number of publications, and quality of department (Wagner, 1987; Wagner and Sternberg, 1985). In studies with salespeople, Wagner et al. (1994) found correlations in the .3 to .4 range between tacit knowledge and criteria such as sales volume and sales awards received. Finally, tacit knowledge for college students was found to correlate with indices of academic performance and adjustment to college (Williams and Sternberg, cited in Sternberg et al., 1993).

In summary, the program of tacit-knowledge research reviewed above shows that generally tacit knowledge increases with experience, but is not simply a proxy for experience; that tacit-knowledge tests measure a distinct construct from that measured by traditional, abstract cognition tests; that scores on tacit-knowledge tests represent a general factor, which appears to correlate across domains; and finally, that tacit knowledge tests are predictive of performance in a number of domains, and compare favorably with those obtained for overall cognitive ability within the range of skills we have tested.

6. An example of the application of the framework: The ALL practical-cognition study

In their everyday lives, people continually need to make situational judgments: how to get along with a difficult boss, how to break bad news to a friend or coworker, how to handle anger or disappointment after a failed endeavor. These skills are important to life adjustment in general, and to workplace efficacy, in particular. This report describes a project aimed at measuring such skills.

The project described here had three major goals. The first goal was to develop a theory-based instrument to measure practical cognition as measured by everyday situational-judgment skills. The second goal was to evaluate the psychometric properties of the instrument, including item characteristics, reliability, and both internal and external validity. The third goal was to compare the psychometric properties and the utility of the instrument in two cultural settings: the United States and Spain.

We also decided that there were certain goals that we explicitly were *not* setting for this project. The first was to measure all possible kinds of situational-judgment skills. In this first phase of our project, we concentrated on workplace situational-judgment skills. In later phases, we will seek to measure other kinds of situational-judgment skillsas well. The second thing we did not try to do was to measure everyday situational-judgment skills in an occupation-specific way. Previously, as described earlier, we had devised a number of inventories for specific occupations, such as managers, salespeople, university professors, university students, military officers, and elementary-school teachers (see Sternberg, Wagner, and Okagaki, 1993; Sternberg, Forsythe, et al., in press; Sternberg, Wagner, Williams, and Horvath, 1995; Wagner and Sternberg, 1986). We intended in this project to extend our methodology to jobs in general rather than devising yet another measure for another specific occupation. Third, we did not seek an inventory with "objectively correct" answers, because situational judgments are, by their nature, more or less useful or possibly justifiable, but they are not, strictly speaking, objectively correct or incorrect. The theoretical basis for our work is the triarchic theory of cognitive skills (Sternberg, 1985a, 1988, 1997, in press-b).

In the current work, we have sought to extend our past work in three ways. First, we have measured informal knowledge that is relatively more domain general than in our past work, where we have targeted specific jobs or career paths. Second, we have sought to extend our findings cross-culturally, using the same inventory in translated form in Spain as in the U.S. Third, we have used item-response-theory (IRT) scaling in order to explore the scalar properties of our inventory.

We also have used a new conceptual framework in this research, which is shown in Figure 1. This framework, used for item construction, crosses three objects of attention (dealing with self, dealing with others, dealing with tasks) with five categories of behavior (motivating, interpreting situations, behaving, following directions, and organizing).

Figure 1
Conceptual framework

	Motivating	Interpreting situations	Behaving	Following directions	Organizing
DS (self)					
DO (others)					
DT (tasks)					

6.1 Method

6.1.1 Materials

Two main kinds of materials were used in this project.

The Everyday Situational Judgment Inventory. The Everyday Situational-Judgment Inventory (ESJI) consists of descriptions of various situations encountered by many people. After each situation, there are 8 options for handling the situation. For each option listed, participants were asked to rate the quality of the option on a 1 (low) to 9 (high) Likert scale, where the anchor points were 1=extremely bad, 3=somewhat bad, 5=neither bad nor good, 7=somewhat good, and 9=extremely good. Participants were asked to select the number corresponding to their judgment, and to write it in the blank preceding each option. Participants were told that there was no one "right" answer—that the options were simply things that people might do in the situations described.

An example of an item is as follows:

You've been assigned to work on a project for a day with a fellow employee whom you really dislike. He is rude, lazy, and rarely does a proper job. What would be the best thing for you to do?

_____*Tell the worker that you think he is worthless.*

_____*Warn the worker that, if he is not "on his toes" today, you will complain to the supervisor.*

_____*Avoid all conversation and eye contact with the other worker.*

_____*Be polite to the other worker and try to maintain as business-like a manner as possible so that hopefully he will follow your example for the day.*

_____*Tell your supervisor that you refuse to work with this man.*

_____*The project is going to be impossible to accomplish with this worker, so you may as well not even try —you can always blame your bad work partner.*

_____*See if you can convince one of your friends to take your place and work with this employee.*

_____*Demand a raise from your supervisor; you should not have to tolerate these conditions.*

Participants were given as much time as they needed to finish the inventory.

Scoring for the ESJI was done in three ways:

1. *Profile matching (d^2).* For each of the 30 problems, a given respondent's responses were compared to the averaged (prototypical) responses to that problem. The following specific scoring procedure was used. For a given option, the difference between the participant's response and the sample-mean response was computed and squared. Squared differences were summed across the 8 options and averaged. Then the square root of this average was computed. The same procedure was repeated for each of the 30 items. Total score was the sum of these values.
2. *Rank-order correlation between individuals and mean profile (ρ).* For this measure, the rank orders of the responses for the mean profile of responses to a given item were correlated with the rank orders of the responses for an individual's profile of responses for that item. Thus, 8 observations for an individual were correlated with 8 observations for the mean profile. This analysis yielded a rank-order correlation (rho, or ρ) for each item. These correlations were averaged across the 30 problems.

3. *Dichotomized responses based on significance of ρ.* Dichotomized scores were created for each of the 30 items by assigning the item a score of 1 if the ρ value for that item was statistically significant and a 0 otherwise. In other words, the respondent got credit (1) if the respondent's item response pattern rank-order correlated significantly with the averaged item response pattern.

Performance evaluations. The performance of the individuals who filled out the ESJI was evaluated in two ways, via self-ratings and via supervisor ratings. In Part 1 of each evaluation (self and supervisor), ratings were coded in the following way: 1=definitely no, 5=not sure, 9=definitely yes. In Part 2, ratings were coded in the following way: 1=extremely bad, 3=somewhat bad, 5=neither bad nor good, 7=somewhat good, 9=extremely good.

Here are the items for self-ratings:

Part 1.

1. *My relationship with my supervisor is good.*
2. *My supervisor thinks highly of me.*
3. *I am satisfied with the development of my career.*
4. *I am planning a career change.*
5. *My relationships with my coworkers are good.*
6. *My greatest strength on the job is my ability to work well with others.*
7. *My greatest strength on the job is my ability to work independently.*
8. *My greatest strength on the job is my ability to manage tasks.*
9. *My greatest strength on the job is to motivate myself.*

Part 2.

1. *How would you rate your common-sense ability?*
2. *How would you rate your academic ability?*
3. *How would you rate your creative ability?*
4. *How good are you at working by yourself?*
5. *How good are you at working with others?*

The supervisors' evaluation rating scales were comparable.

Part 1.

1. *My relationship with this employee is good.*
2. *I think highly of this employee.*
3. *I am satisfied with this employee.*
4. *The employee's relationships with other coworkers are good.*

Part 2.

1. *How would you rate this employee's common sense ability?*
2. *How would you rate this employee's academic ability?*
3. *How would rate this employee's creative ability?*
4. *How would you rate this employee at working by him/herself?*
5. *How would you rate this employee at working with others?*
6. *How good is this employee at motivating him/herself?*
7. *How good is this employee at managing tasks?*
8. *How responsible is this employee?*

6.1.2 Participants

There were two sets of participants, from the United States and from Spain.

U.S. participants. There were 230 U.S. participants, 78 male, 149 female, 3 unidentified. The mean age was 35.8 years with a standard deviation of 13.5 years. The range was from 17 to 72 years of age. Mean time in the workplace was 6.7 years with a standard deviation of 7.9 years. Mean time in current position was 1.3 years with a standard deviation of 1.0 years. Job classifications of these participants included custodians, food-service workers in a dining hall, restaurant waiters and waitresses, salespeople, postal-service workers, taxi drivers, office personnel, and teachers.

Spanish participants. There were 227 Spanish participants, 112 male, 112 female, and 3 unidentified. The mean age was 36.1 years with a standard deviation of 9.8 years. The range was from 21 to 64 years of age. Mean time in the workplace was 7.6 years with a standard deviation of 8.6 years. Mean time in current position was 4.0 years with a standard deviation of 3.6 years. Job classifications of these participants clerks, bank office personnel, photography studio personnel, biology laboratory personnel, film developing studio personnel, lawyers' office support personnel, librarians, educational researchers, textbook editors, university teachers, air traffic controllers, administrative personnel of diverse institutions, and psychiatrists.

6.1.3 Design

The main dependent variables were responses to the performance-evaluation items (as answered by both workers and their supervisors). The main independent variables were scores on the ESJI.

6.1.4 Procedure

The ESJI administered individually in both the United States and Spain. The administration typically took 30-40 minutes, although the inventory was untimed. The instrument was constructed in the United States and then translated into Castilian Spanish by the Spanish team and checked for accuracy by a Spanish-speaking member of the U.S. team.

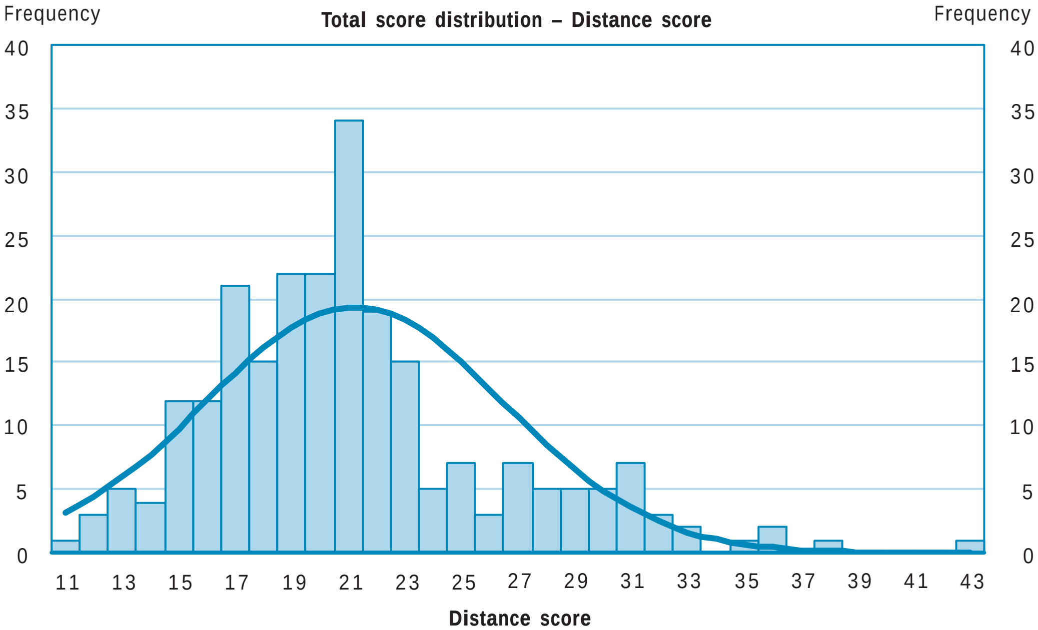
6.2 Results and discussion

6.2.1 Basic statistics and score distributions

Figure 2 (Panels A-F) shows the total-score distributions for each of the three methods of scoring for each of the samples (United States and Spain). Each figure also shows the mean and standard deviation for the given scoring system for the given sample, as well as the N on which the set of sample statistics is based.

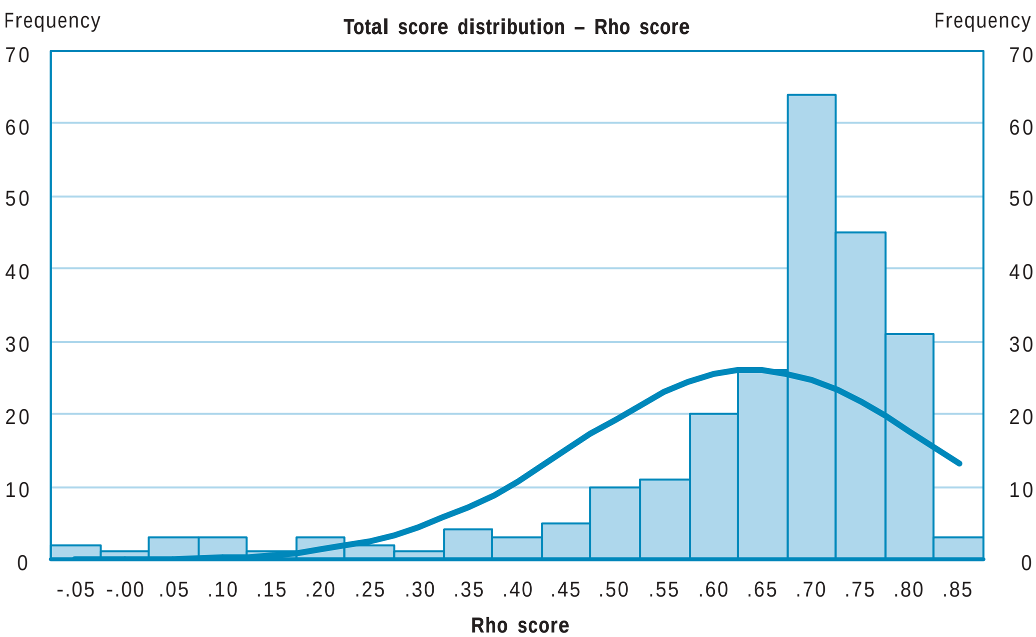
The Spanish sample performed better than did the U.S. sample, although because the occupations of the Spanish and U.S. samples were not exactly equated, the samples may not have been strictly comparable. The differences were statistically significant at the .001 level, regardless of the measure used. For the distance scores, t_{455} was equal to 8.47. For the rank-order correlation scores, the comparable t-value was 5.32, and for the 1/0 scores, 5.92.

Figure 2 (Panel A)
Total score characteristics: US sample



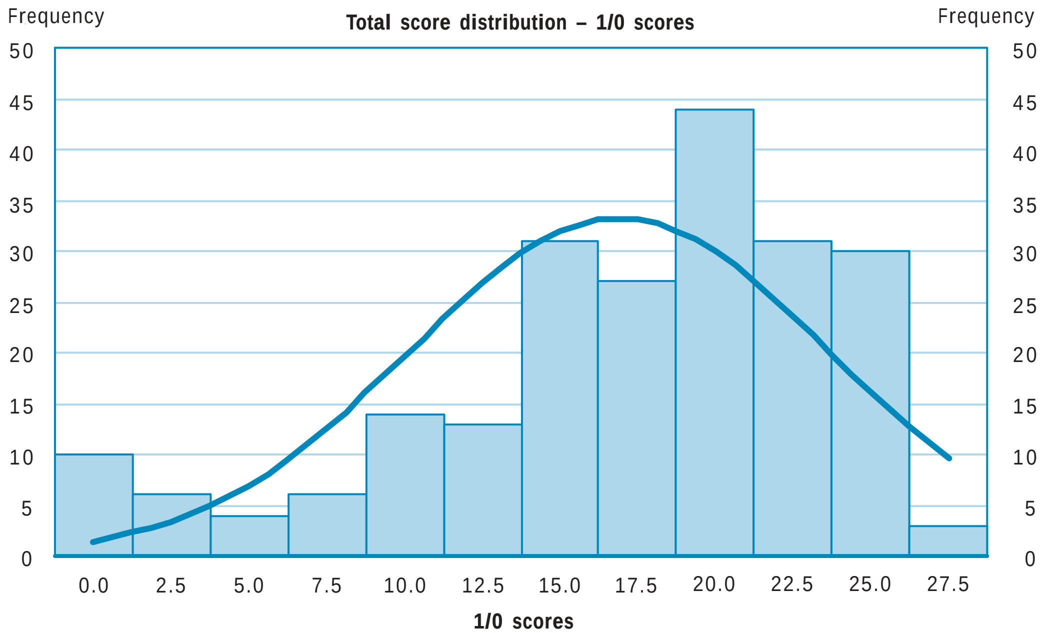
Standard deviation = 5.31
Mean = 21.2
Number = 230.00

Figure 2 (Panel B)
Total score characteristics: US sample



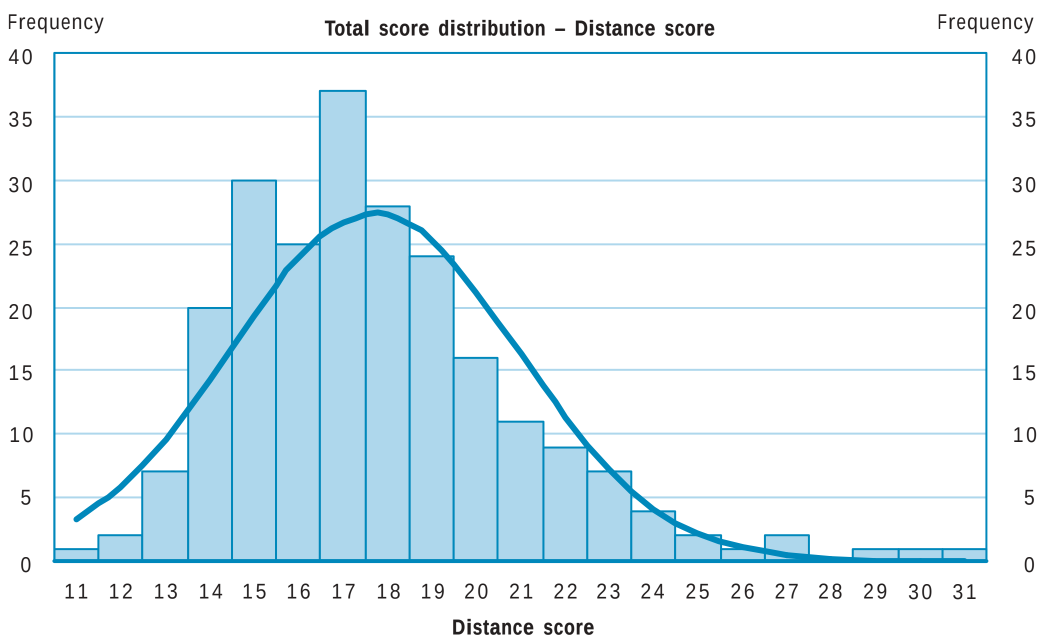
Standard deviation = .18
Mean = .64
Number = 230.00

Figure 2 (Panel C)
Total score characteristics: US sample



Standard deviation = 6.70
Mean = 16.9
Number = 230.00

Figure 2 (Panel D)
Total score characteristics: Spanish sample



Standard deviation = 3.24
Mean = 17.7
Number = 227.00

Figure 2 (Panel E)
Total score characteristics: Spanish sample

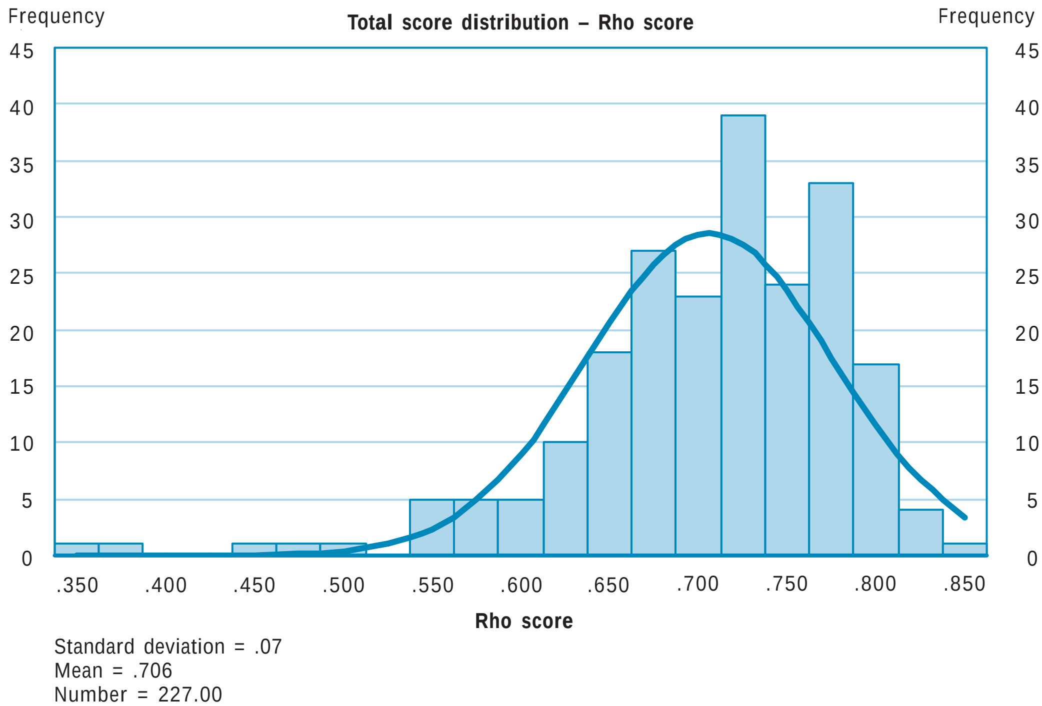
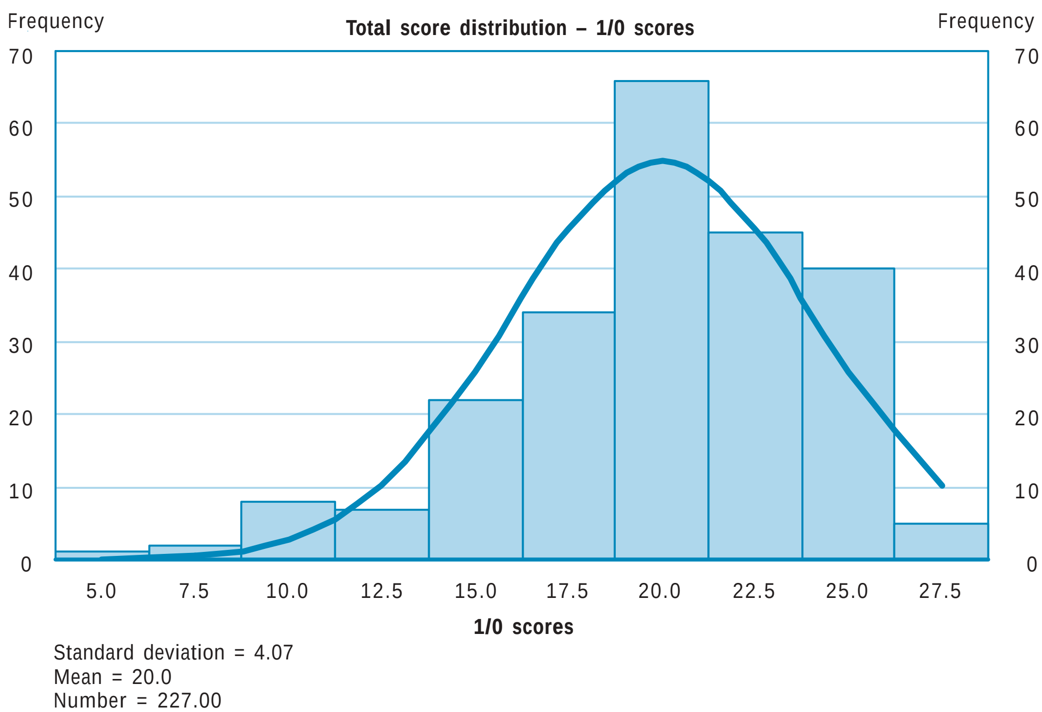


Figure 2 (Panel F)
Total score characteristics: Spanish sample



6.3 Internal validation

6.3.1 Internal-consistency reliabilities

For distance scores, coefficient a internal-consistency reliabilities were .96 for the U.S. sample and .92 for the Spanish sample. For rank-order correlation scores, internal-consistency reliabilities were .94 for the U.S. sample and .73 for the Spanish sample. For 1/0 scores, KR-20 internal-consistency reliabilities were .89 for the U.S. sample and .62 for the Spanish sample. The overall 1/0 internal-consistency reliability for the combined samples was .82.

6.3.2 Comparison of responses between U.S. and Spanish samples

The correlation for mean profiles of responses to item options was r = .91, and the correlation for standard deviations of responses to item options was r = .66 (with 8 options for 30 items, or 240 observations, in each set). These correlations indicate that the responses across countries were about as similar as one could hope for, given the reliabilities of the data.

6.3.3 Item characteristics

We analyzed item characteristics for the combined samples using 1/0 scoring. The range of facilities (p-values) was from .42 to .84, with a mean p-value of .62. The range of difficulties (Δ values) was from 9.1 to 13.4 with a mean of 11.8. The range of discriminating power for items was computed with both biserial and point-biserial correlations. The range of r_{bis} was from .38 to .72, with a mean of .53. The range of r_{pbis} was from .29 to .53, with a mean of .41.

Specifics for the 1-Parameter Logistic (Rasch) and 2-Parameter Logistic (Birnbaum) Models. One- and two-parameter-logistic (PL) models, as derived from item-response theory (IRT), were fit to the data. The difference in the maximal marginal likelihoods was 84.7, with 30 degrees of freedom, indicating the superior fit of the two-parameter model. For the one-parameter model, 7 items had statistically significant chi-squares ($p < .05$), whereas for the two-parameter model, none of the items did.

6.3.4 Item-information statistics

Table 1 shows the point of maximum information for each of the 30 items in the ESJI. The results suggest that the items tended to be rather easy, on the whole, and that subsequent versions of the ESJI probably need some more difficult items.

Table 1

Item information statistics

Point of Maximum Information

1.	- .5335	11.	- .3486	21.	-1.3415
2.	-1.8624	12.	- .2408	22.	.0778
3.	- .0595	13.	-1.1276	23.	- .8488
4.	- .1422	14.	-1.4116	24.	-1.1307
5.	- .3898	15.	-1.4472	25.	-1.3093
6.	-1.5271	16.	- .8430	26.	- .1914
7.	.0279	17.	- .6911	27.	- .6369
8.	-1.1277	18.	.1810	28.	- .6239
9.	-1.7692	19.	.2232	29.	- .7331
10.	- .2972	20.	.5614	30.	.3627

6.3.5 Factor analyses

Exploratory factor analysis. Exploratory principal-factor analysis was done, yielding 5 factors with eigenvalues greater than 1 in the Spanish sample (and .9 in the U.S. sample). These factors accounted for roughly 50% of the variance in the Spanish data and 63% of the variance in the U.S. data. Eigenvalues were 9.9, 1.5, 1.2, 1.2, and 1.1 respectively for the Spanish data and 14.3, 1.5, 1.2, 1.0, and .9 in respectively for the U.S. data.

Confirmatory factor analysis. Confirmatory factor analysis also was done. The conceptual framework proposed in Figure 1 provided a 3 x 5 model for the items of the ESJI. We therefore evaluated corresponding three- and five-factor latent variable models for the U.S. and Spanish data.

Initially, we evaluated 4 different models: (a) a 3-factor model for the U.S. sample, $\bullet_{(402)}^{2}$ = 918.4, RMSEA (root-mean-square error of approximation) = .08, CFI (comparative fit index) = .87, IFI (incremental fit index) =.87, and GFI (goodness of fit index) = .78; (b) a 3-factor model for the Spanish sample, $\bullet_{(402)}^{2}$ = 582.2, RMSEA = .043, CFI = .91, IFI = .91, and GFI = .86; (c) a 5-factor model for the USA sample, $\bullet\ (_{395)}^{2}$ = 878.6, RMSEA = .08, CFI = .88, IFI =.88, and GFI =.79; and (d) a 5-factor model for the Spanish sample, $\bullet_{(395)}^{2}$ = 526.0, RMSEA = .04, CFI = .94, IFI = .94, and GFI = .87. Two conclusions were drawn from these results. First, the model-fit indexes were comparable for the U.S. and Spanish samples, suggesting that the data could be combined in a single analysis. Second, overall, the fit indexes were better for the 5-factor model than for the 3-factor model, suggesting that the 5-factor structure was the preferred latent structure of the inventory. Because results were comparable for the U.S. and Spanish samples, we combined them in a single 5-factor multi-group model. We fitted four different modifications of this model: (a) a model equating the correlations between the latent variables in both samples, $\bullet_{(800)}^{2}$ = 1421.2, RMSEA = .06, CFI = .90, IFI = .90, and GFI = .87; (b) a model equating the correlations between the latent variables and measurement errors in both samples, $\bullet_{(830)}^{2}$ = 1643.6, RMSEA = .07, CFI = .87, IFI = .87, and GFI = .83; (c) a model equating the correlations between latent variables and factor loadings of the measured variables on the latent variables, $\bullet_{(830)}^{2}$ = 1627.6, RMSEA = .07, CFI = .87, IFI = .87, and GFI = .83; and (d) a model equating the correlations between the latent variables, measurement errors, and factors loading of the observed variables on the latent variables, •$_{(860)}^{2}$ = 1857.5, RMSEA = .08, CFI = .84, IFI = .84, and GFI = .80. According to these indexes, Model (a) described the data the best, suggesting that the underlying latent structure of the inventory is invariant across the U.S. and Spanish samples, but the measurement errors and factor loadings differ in the two samples.

The variables in the five-factor model were related. Table 2 shows the intercorrelations of the latent variables. As can be seen in the table, these latent-variable correlations are extremely high, suggesting that the five factors of the model are highly correlated and may represent a general factor, although not necessarily psychometric *g*, given that in past research subscales also have been highly correlated with each other but not with psychometric *g*.

Table 2

Latent variable correlations

	M	S	B	D	O
Motivating	1.00				
Situations	0.91	1.00			
	(0.02)				
	41.29				
Behaving	0.87	0.95	1.00		
	(0.02)	(0.02)			
	36.03	52.79			
Directions	0.84	0.94	0.96	1.00	
	(0.03)	(0.02)	(0.02)		
	32.25	50.09	56.77		
Organizing	0.87	0.95	0.98	0.96	1.00
	(0.02)	(0.02)	(0.02)	(0.02)	
	35.61	53.67	63.20	57.26	

6.4 External validation

6.4.1 Concurrent validities

To stress the consistency of the results across methods of scoring, we show r-based validity coefficients for the sample-specific data and 2-PL rescaled r-based validity coefficients for the combined sample data.

Supervisors' ratings. Concurrent validities for the rank-order correlation scoring of the ESJI are shown in Table 3 for the U.S. sample, the Spanish sample, and both samples combined. For the U.S. sample, validity coefficients ranged from .22 to .46 with a median of .36. All (8 of 12) correlations were statistically significant at the .05 level. For the Spanish sample, validity coefficients ranged from -.10 to .21 with a median of .14. Only three of the correlations were statistically significant. For the total sample, validity coefficients based on the 2-PL model ranged from .09 to .32 with a median of .20. Ten of the 12 correlations were statistically significant. Correlations were approximately the same for the d^2 scores and actually slightly better with the 1-PL model.

Table 3
Concurrent validities

Spanish sample (rhos)		
1.	My relationship with this employee is good:	-.10
2.	I think highly of this employee:	.10
3.	I am satisfied with this employee:	.05
4.	The employees relationships with other coworkers are good:	.06
1.	…Common sense ability:	.21 *
2.	…Academic ability:	.21 *
3.	…Creative ability:	.13
4.	…Working by him/herself:	.07
5.	…Working with others:	.14
6.	…Motivating him/herself:	.02
7.	…Managing tasks:	.16
8.	…Responsible:	.20 *
US sample (rhos)		
1.	My relationship with this employee is good:	.26 **
2.	I think highly of this employee:	.34 ***
3.	I am satisfied with this employee:	.38 ***
4.	The employees relationships with other coworkers are good:	.22 *
1.	…Common sense ability:	.42 ***
2.	…Academic ability:	.44 ***
3.	…Creative ability:	.37 ***
4.	…Working by him/herself:	.34 ***
5.	…Working with others:	.29 **
6.	…Motivating him/herself:	.46 ***
7.	…Managing tasks:	.41 ***
8.	…Responsible:	.30 **
Total sample (rescaled, 2-PL [Birnbaum] model)		
1.	My relationship with this employee is good:	.12
2.	I think highly of this employee:	.18 **
3.	I am satisfied with this employee:	.21 **
4.	The employees relationships with other coworkers are good:	.09
1.	…Common sense ability:	.32 ***
2.	…Academic ability:	.31 ***
3.	…Creative ability:	.20 **
4.	…Working by him/herself:	.17 *
5.	…Working with others:	.17 *
6.	…Motivating him/herself:	.18 **
7.	…Managing tasks:	.25 ***
8.	…Responsible:	.20 **

Note: Correlations are slightly better with 1PL

We cannot say for sure why the correlations for the U.S. sample were better than those for the Spanish sample. Some possible interpretations, based on what we know of the data, are (a) differential quality of supervisory ratings, (b) differential compositions of the two respective samples, and (c) relatively greater homogeneity of the Spanish sample relative to the U.S. sample. In particular, Spanish supervisors seemed less comfortable providing ratings of quality of performance than did U.S. supervisors, but because we did not anticipate this result, we collected no quantitative data with respect to it.

Self-ratings. The distributions of self-ratings proved not to be amenable to the validity analyses due to restriction of range in these variables. As has been found in other studies (e.g., Sternberg et al., 1981), participants tend to have almost a uniformly high opinion of themselves. With the item on planning career change deleted (because this item does not measure self-evaluation of competence), mean self-ratings on the 9-point scale ranged from 6.5 to 8.3, with an overall mean of 7.7 (sd = .47) for the United States sample and from 6.7 to 7.8 for the Spanish sample, with an overall mean of 7.1 (sd = .39). Corresponding standard deviations ranged from 1.0 to 2.4, with an overall mean standard deviation of 1.4 (sd = .40) for the United States sample. For the Spanish sample, standard deviations ranged from 1.1 to 1.9 with an overall mean of 1.5 (sd = .30). Ratings also were highly leptokurtic. Given the problem of restriction of range, we did not pursue further analyses of the self-ratings.

In sum, our main findings were of (a) satisfactory psychometric properties for the ESJI, (b) satisfactory to excellent internal-consistency reliabilities of the inventory, (c) excellent consistency of the data across cultures, (d) satisfactory internal validity using five-factor model, and (e) moderate concurrent validity in the U.S. sample and modest concurrent validity in the Spanish sample. We believe that the study showed the feasibility of our approach in measuring practical cognition.

7. Conclusions

Approximately 25 years ago, McClelland (1973) questioned the validity of cognitive-skill testing for predicting real-world criteria such as job performance, arguing in favor of competency tests that more closely would reflect job performance itself. Subsequent reviews of the literature on the predictive validity of cognition tests suggest that McClelland may have been pessimistic about the validity of cognition tests: individual differences in cognition-test performance account for, on average, between 4 and 25 percent of the variance in real-world criteria such as job performance (Barrett and Depinet, 1991; Hunter and Hunter, 1984; Schmidt and Hunter, 1998; Wigdor and Garner,1982). Nevertheless, these findings indicate that between 75 and 96 percent of the variance in real-world criteria such as job performance cannot be accounted for by individual differences in cognition-test scores. The emerging literature on practical cognition and similar constructs (e.g., social and emotional cognition) is a belated response to McClelland's call for new methods to assess practical skills. The literature and research reviewed in this volume provide several sources of evidence to support a distinction between academic and practical cognition.

First, the distinction between academic and practical cognition is entrenched in the conception of cognition held by laypeople and researchers alike. In addition to evidence provided by studies of implicit theories of cognition (e.g., Sternberg et al., 1981), analyses of researchers' descriptions of the nature of cognition suggest a prominent role for practical cognition. Seventy years ago, the editors of the *Journal of Educational Psychology* convened a symposium at which prominent psychological theorists of the day were asked to describe what they imagined cognition to be and what they considered the most crucial "next steps" in research. In a replication, Sternberg and Detterman (1986) posed these same questions to contemporary prominent theorists. An analysis of the responses of both cohorts of cognition theorists revealed concerns about practical aspects of cognition (Sternberg and Berg, 1986). For example, among the 42 crucial next steps that were mentioned by one or more theorists from either cohort, studying real-life manifestations of cognition was among the most frequently mentioned "next steps" of both the contemporary researchers and the original respondents. A distinction between academic and practical aspects of cognition also is supported by older adults' perception of age-related changes in their skill to think and solve problems (Williams, Denney, and Schadler, 1983). Three-fourths of the older adults sampled believed that their skill to solve practical problems increased over the years, despite the fact that performance on academic tasks begins to decline upon completion of formal schooling.

A second source of evidence to support a distinction between academic and practical cognition is the set of results of studies in which participants were assessed on both academic and practical tasks. These studies consistently find little or no correlation between performance on the two kinds of tasks. Overall, cognitive ability tests and similar measures are unrelated to (a) the order-filling performance of milk-processing plant workers (Scribner, 1986); (b) the degree to which racetrack handicappers employ a complex and effective algorithm (Ceci and Liker, 1986, 1988); (c) the complexity of strategies used in computer-simulated roles such as city manager (Dörner and Kreuzig, 1983; Dörner et al., 1983); and (d) the accuracy with which grocery shoppers identified quantities that provided the best value (Lave et al., 1984; Murtaugh, 1985). This research shows that the performance of both children and adults is susceptible to the context in which skills are measured. When problems are presented in a familiar context, whether that context is school or work, individuals appear more intelligent (e.g., Carraher et al., 1985; Roazzi, 1987).

A third source of support for the importance of practical skills comes from theories of managerial performance. Rational theories that are based on conventional notions of how people solve problems (e.g., Kepner and Tregoe, 1965; Plunkett and Hale, 1982) do not accurately represent the problem solving of experienced and successful managers. These observations led theorists to describe managerial problem solving as non-linear, convoluted, and action-oriented (e.g. McCall and Kaplan, 1985; Mintzberg et al., 1976). Furthermore, knowledge of how to solve problems can be characterized as tacit, and it may only enter into conscious awareness through reflection (Schön, 1983). The recognition that rational models of managerial problem solving do not explain the behavior of successful practitioners suggests that alternative approaches are needed to identify the practical skills underlying performance.

Finally, the research on tacit knowledge described throughout this volume offers an approach to understanding practical cognition. Over the course of studies with academic psychologist (Wagner, 1987; Wagner and Sternberg, 1985), business managers (Wagner and Sternberg, 1990), salespersons (Wagner et al., 1994), U.S. Air Force recruits (Eddy, 1988), and military leaders (Hedlund et al., 1999), we have found that tacit knowledge offers insight into the practical skills associated with success.

Several conclusions can be drawn from this program of research. First, these studies showed that tacit knowledge exists in the stories successful practitioners share about the lessons they learned in the process of performing their respective roles. These stories provide rich insights about the practically-oriented knowledge that practitioners are often unaware that they have acquired. Second, we showed that tacit knowledge can be measured through instruments that take into account the procedural and context-specific nature of tacit knowledge. Third, using such instruments, we have found that individuals who exhibit the skill to acquire and use tacit knowledge are more effective in their respective performance domains. Furthermore, tacit knowledge helps to explain some of the additional variance in performance that is not accounted for by measures of general cognitive skill. Fifth, although the acquisition of tacit knowledge may be influenced, to some extent, by *g* and amount of experience, tacit-knowledge inventories are not simply new measures of these constructs. Finally, tacit knowledge generally appears to be a singular construct within domains, but the content of tacit knowledge varies across domains. In other words, tacit knowledge appears to reflect a single underlying skill, which we label practical cognition. But, this underlying skill is not sufficient for performing well on domain-specific tacit-knowledge tests. Experience in a particular domain is important in the acquisition of tacit knowledge.

Based on consistent findings that tacit knowledge contributes to our understanding performance in a variety of domains, we discussed a number of potential ways to promote the acquisition and use of tacit knowledge. Numerous insights and products are obtained through the process of studying tacit knowledge. The categories of tacit knowledge within a domain, for example, offer insight into the experiences that provide important developmental opportunities. The products, such as the stories and the inventory questions, can be used to share the tacit knowledge with other practitioners. The tacit-knowledge research also suggests that training activities, such as case studies and simulations, may be valuable ways to impart experience-based, tacit knowledge and to provide opportunities to acquire new practical knowledge. Although these approaches may encourage the acquisition and use of tacit knowledge, in rapidly changing, complex environments, it may be more effective in the long run to identify and develop ways to help individuals to learn better from their everyday experiences.

Up to this point, our research efforts have been targeted primarily at understanding and measuring practical cognition. For the present and foreseeable future, we believe that the most viable approach to increasing the variance accounted for in real-world criteria such as job performance is to supplement existing cognition and aptitude tests with selection of additional measures based on new constructs such as practical cognition. Although we are excited by the promise of a new generation of measures of practical cognition, we are the first to admit that existing evidence for the new measures does not yet match that available for traditional cognitive-academic skill tests. However, a substantial amount of evidence indicates that performance on measures of practical cognition is related to a wide variety of criterion measures of real-world performance, but relatively unrelated to traditional measures of academic cognition. Consequently, using both kinds of measures explains more variance in performance than relying on either kind alone. Cognition is not only academic, but practical.

8. Future directions

There still are some issues that need to be addressed in further work on the scale, and we plan to address these issues in the next version.

1. Face validity. The use of the term *boss* proved to be a mistake, as it carries a somewhat negative connotation in European and other countries were hierarchical arrangements of workers are less socially acceptable than in the United States. The inventory also relied too much on office-type settings and needs to be expanded to include not only a broader range of occupational settings, but settings outside of the workplace. Our goal in the next version of our questionnaire, therefore, is to do extensive revision for face validity to ensure that the questionnaire we use will be viewed as face valid by all test-takers. Following a procedure we have used in other research, our plan is to ask test-takers to evaluate the face validity of the questionnaire by asking them how realistic the scenarios and solutions are.
2. Length. The ESJI was longer than would be ideal, given realistic constraints upon administration time. It could be and should be shortened both with respect to number of items (from 30 to, perhaps, 20) and number of items (from 8 to, perhaps, 5).
3. Description of test. The test originally was described as a test of practical ability or of everyday cognition. We have changed the name, effective as of this article, to reflect better both what the test measures and the constraints of the sociopolitical context in which the test will be administered. We thus will refer to the test as an everyday-situation-judgment test.
4. Number of scale points. The number of scale points were item will be seven in order to ensure that each point supplies useful information. Each scale point will have a verbal label describing its meaning.
5. Need for anchors in scoring. Scoring in the future will be converted to a 0-500 scale in order to match other assessments in the study of adult competencies in OECD countries. The mean will be 250 and the standard deviation, 50. We will anchor score points to specific levels of competency verbally described. Specific levels of competency will be anchored to various indicators of job performance, career satisfaction, and life satisfaction (with relevant data to be obtained from a new validity study).
6. Objective scoring. In addition to the type of prototype scoring we have used, we will experiment with right-wrong scoring based on sets of values that our previous research indicates seem to be universally accepted as indicating preferred behavior (e.g., honesty, sincerity, hard work, compassion, etc.)
7. Uninformative options. Options that elicit means very near the middle of the scale together with high standard deviations will be eliminated, as such options tend to be ones on which there is substantial disagreement among respondents.
8. Skills taxonomy. We already have begun development of a skills taxonomy and this development will be formalized for the next version of the ESJI. We have discovered that certain behaviors tend to be valued cross-situationally and other behaviors to be devalued. Examples of valued behaviors are productivity, honesty, politeness, serving as a good example to others, and doing what needs to be done even if one does not want to do it. Examples of devalued behaviors are lying, not working when one should, passing the buck, blaming others for one's own mistakes, failing to meet one's responsibilities, and doing things grudgingly.
9. Test-retest reliabilities. The present design did not allow for computation of test-retest reliabilities to measure scale stability, but future studies will do so.
10. Discriminant validities. We need a subsequent design that will permit us to assess the discriminant validity of the test with respect to *g*-based measures of abilities.
11. Range of skills tested. We need to develop a wider range of behaviors to be sampled, including behaviors from more varied jobs and behaviors that occur outside of jobs.

12. Translation. The procedures for translation were rather informal in this study and resulted in a few items that, in retrospect, were not ideally translated. The result was a lowering of the cross-sample correlation (which was nevertheless .91). A more careful procedure for ensuring the accuracy of translation is needed in a subsequent study, as well as for ensuring that the situations sampled adequately represent apposite situations in the two cultures assessed as well as others.
13. Focus groups. In subsequent work, we also plan prior to validation to have focus groups in each culture that will evaluate the relevance of each potential inventory situation for cultural appropriateness. Scenarios that are not deemed to be culturally valid by two-thirds of the focus group will be replaced.
14. Occupational groups. We will score by occupational groups in order to determine whether different occupational groups tend to prefer different responses to options.
15. Cultures. We would hope to introduce a third cultural context in subsequent work. Sternberg, Grigorenko, and Gil are all willing to participate in subsequent phases of this work.

In sum, the ESJI appears to be a promising inventory for the assessment of situational judgments, but further work is needed in order to refine the existing measure.

References

Ackerman, P. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27.

Ackerman, P. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D.K. Detterman (Ed.), *Current topics in human intelligence: Theories of cognition* (Vol. 4, pp. 1-27). Norwood, NJ: Ablex.

Ackerman, P.L., and Heggestad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219-245.

Aldwin, C.M., Sutton, K.J., Chiara, G., and Spiro, A. (1996). Age differences in stress, coping, and appraisal: Findings from the normative aging study. *Journal of Gerontology: Psychological Sciences, 51B*, 178-188.

Anastasi, A., and Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J.R. (1986). Knowledge compilation: The general learning mechanism. In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach.* Los Altos, CA: Kaufman.

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Anderson, J.R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Archer, D. (1980). *How to expand your social intelligence quotient.* New York: M. Evans.

Archer, D., and Akert, R.M. (1980). The encoding of meaning: A test of three theories of social interaction. *Sociological Inquiry, 50,* 393-419.

Argyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review, 69,* 99-109.

Bäckman, L. and Dixon, R.A. (1992). Psychological compensation: A theoretical framework. *Psychological Bulletin, 112,* 259-283.

Baltes, P.B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology, 23*, 611-626.

Baltes, P.B. (1993). The aging mind: Potentials and limits. *The Gerontologist, 33,* 580-594.

Baltes, P.B. (1997). On the incomplete architecture of human ontogeny. *American Psychologist, 52*, 366-380.

Baltes, P. (in press). *Wisdom: the orchestration of mind and virtue.* Boston: Blackwell.

Baltes, P.B. and Baltes, M.M. (1990). Psychological perspectives on successful aging: A model of selective optimization with compensation. In P.B. Baltes and M.M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences.* Cambridge: Cambridge University Press.

Baltes, P.B., Dittmann-Kohli, F., and Dixon, R.A. (1984). New perspectives on the development of intelligence in adulthood: Toward a dual-process conception and a model of selective optimization with compensation. In P.B. Baltes and O.G. Brim (Eds.), *Life-span development and behavior* (Vol 6., pp. 33-76). New York: Academic Press.

Baltes, P.B., and Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy, 19,* 282-300.

Baltes, P.B., Sowarka, D., and Kliegl, R. (1989). Cognitive training research on fluid intelligence in old age: What can older adults achieve by themselves? *Psychology and Aging, 4,* 217-221.

Baltes, P.B., and Willis, S.L. (1982). Plasticity and enhancement of cognitive functioning in old age: Penn State's Adult Development and Enrichment Project (ADEPT). In F.I.M. Clark and S.E. Trehub (Eds.), *Aging and cognitive processes* (pp. 353-389). New York: Plenum.

Band, E. B., and Weisz, J.R. (1988). How to feel better when it feels bad: Children's perspective on coping with everyday stress. *Developmental Psychology, 24,* 247-253.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs, NJ: Prentice Hall.

Barker, R.G. (Ed.) (1978). *Habitats, environments, and human behavior.* San Francisco: Jossey-Bass.

Barnes, M.L., and Sternberg, R.J. (1989) Social intelligence and decoding of nonverbal clues. Intelligence, *13,* 263-287.

Bar-On, R. (1997). *The Emotional Quotient Inventory (EQI): Technical Manual.* Toronto, Canada: Multi-Health Systems.

Barrett, G.V., and Depinet, R.L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46,* 1012-1024.

Barrett, G.V., Mihal, W.L., Panek, P.E., Sterns, H.L., and Alexander, R.A. (1977). Information processing abilities predictive of accident involvement for younger and older commercial drivers. *Industrial Gerontology, 4,* 173-182.

Barrett, G.V., and Watkins, S.K., (1986). Word familiarity and cardiovascular health as determinants of age-related recall differences. *Journal of Gerontology, 41,* 222-224.

Beck, A.T., Epstein, N., Brown, G., and Steer, R.A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893-897.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Belmont, J.N., and Butterfield, E.C. (1969). The relations of short-term memory to development and cognition. In L. Lipsitt and H. Reese (Eds.), *Advances in child development and behavior* (Vol. 4., pp. 30-83). New York: Academic Press.

Bennett, G.K., Seashore, H.G., and Wesman, A.G. (1974). *The Differential Aptitude Tests* (Form T). New York: The Psychological Corporation.

Berg, C.A. (1984). Perspectives for viewing cognitive development throughout the life course. In R.J. Sternberg and C.A. Berg (Eds.), *Cognitive Development* (pp. 1-15). New York, NY: Cambridge University Press.

Berg, C.A. (1989). Knowledge of strategies for dealing with everyday problems from childhood through adolescence. *Developmental Psychology, 25,* 607-618.

Berg, C.A. (in press). The development of adult intelligence. In R.J. Sternberg (Ed.), *Handbook of intelligence*. New York, NY: Cambridge University Press.

Berg, C.A., and Calderone, K. (1994). The role of problem interpretations in understanding the development of everyday problem solving. In R.J. Sternberg and R.K. Wagner (Eds.), *Mind in context* (pp. 105-132). New York, NY: Cambridge University Press.

Berg, C.A., Calderone, K. and Gunderson, M. (1990, November). *Strategies young and old adults use to solve their own everyday problems.* Paper presented at the meeting of the Gerontological Society, Boston, MA.

Berg, C.A., Calderone, K., Sansone, C., Strough, J., and Weir, C. (1998). The role of problem definitions in understanding age and context effects on strategies for solving everyday problems. *Psychology and Aging, 13,* 29-44.

Berg, C.A., Hertzog, C., and Hunt, E. (1982). Age differences in the speed of mental rotation. *Developmental Psychology, 18,* 95-107.

Berg, C.A., Klaczynski, P. (1996). Practical intelligence and problem solving: Searching for perspective. In F. Blanchard-Fields and T.M. Hess (Eds.), *Perspectives on intelligence in adulthood and aging* (pp. 323-357). New York: McGraw-Hill.

Berg, C.A., Klaczynski, P., Calderone, K.S., and Strough, J. (1994). Adult age differences in cognitive strategies: Adaptive or deficient. In J. Sinnott (Ed.), *Interdisciplinary Handbook of Adult Lifespan Learning* (pp. 371-388). Westport, CT: Greenwood Press.

Berg, C.A., and Sternberg, R.J. (1985). A triarchic theory of cognitive development during adulthood. *Developmental Review, 5,* 334-370.

Bernadin, H.J., and Smith, P.C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS). *Journal of Applied Psychology, 65,* 458-463.

Berry, J.W. (1974). Radical cultural relativism and the concept of intelligence. In J.W. Berry and P.R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural psychology* (pp. 225-229). London: Methuen.

Berry, J.W. (1984). Towards a universal psychology of cognitive competence. In P.S. Fry (Ed.), *Changing conceptions of cognition and cognitive functioning* (pp. 35-61). Amsterdam: North-Holland.

Berry, J.W., and Irvine, S.H. (1986). Bricolage: Savages do it daily. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 271-306). New York: Cambridge University Press.

Bickhard, M.H. (1978). The nature of developmental stage. *Human Development, 21,* 217-233.

Blanchard-Fields, F. (1986). Reasoning and social dilemmas varying in emotional saliency: An adult developmental perspective. *Psychology and Aging, 1,* 325-333.

Blanchard-Fields, F. (1994). Age differences in causal attributions from an adult developmental perspective. *Journal of Gerontology: Psychological Sciences, 49*, 43-51.

Blanchard-Fields, F., Jahnke, H.C., and Camp, C. (1995). Age differences in problem-solving style: The role of emotional salience. *Psychology and Aging, 10,* 173-180.

Blanchard-Fields, F., and Norris, L. (1994). Causal attributions from adolescence through adulthood: Age differences, ego level, and generalized response style. *Aging Neuropsychology and Cognition, 1,* 67-86.

Bloom, B.S. (Ed.) (1985). *Developing talent in young people.* New York: Ballantine.

Bogen, J.E. (1975). Some educational aspects of hemispheric specialization. *UCLA Educator, 17,* 24-32.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol. 2* (pp. 271-326). Palo Alto, CA: Consulting Psychologists Press, Inc.

Boyatzis, R.E. (1982). *The competent manager*. New York: John Wiley.

Brandtstaedter, J., and Greve, W. (1994). The aging self: Stabilizing and protective processes. *Developmental Review, 14,* 52-80.

Bray, D.W. (1982). The Assessment Center and the study of lives. *American Psychologist, 37,* 180-189.

Broadbent, D.E., and Aston, B. (1978). Human control of a simulated economic system. *Ergonomics, 21,* 1035-1043.

Broadbent, D.E., Fitzgerald, P., and Broadbent, M.H.P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77,* 33-50.

Broca, P.P. (1861). Nouvelle observation d'aphemie produite par une lésion de la moitié postérieure des deuxième et troisième circonvolutions frontales gauches. *Bulletins de la Societé Anotomique de Paris, 36,* 398-407.

Bronfenbrenner, U., and Ceci, S.J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101,* 568-586.

Brown, A.L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 103-152). New York: Academic Press.

Brown, L.T., and Anthony, R.G. (1990). Continuing the search for social intelligence. *Personality and Individual Differences, 11,* 463-470.

Bruner, J.S., Shapiro, D., and Tagiuri, R. (1958). The meaning of traits in isolation and in combination. In R. Tagiuri and I. Petrollo (Eds.), *Person perception and interpersonal behavior* (pp. 277-288). Stanford, CA: Stanford University Press.

Burke, M.J., and Day, R.R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71,* 232-246.

Campbell, D.T, and Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Campbell, J.P. (1988). Training design for performance improvement. In J.P. Campbell, R.J. Campbell, and Associates (Eds.), *Productivity in organizations* (pp. 177-216). San Francisco: Jossey-Bass.

Cantor, N. (1978). Prototypicality and personality judgments. Unpublished doctoral dissertation, Stanford University.

Cantor, N. (1990). From thought to behavior: "Having" and "doing" in the study of personality and intelligence. *American Psychologist*, *45*, 735-750.

Cantor, N., and Harlow, R. (1994). Social intelligence and personality: Flexible life-task pursuit. In R.J. Sternberg and P. Ruzgis (Eds.), *Personality and intelligence* (pp. 137-168). Cambridge, U.K.: Cambridge University Press.

Cantor, N. and Kihlstrom, J.F. (1987). *Personality and social intelligence.* Englewood Cliffs, N.J.: Prentice Hall.

Cantor, N., and Kihlstrom, J.F. (1989). Social intelligence and intellectual assessments of personality. In R.S. Wyer and T.K. Srull (Eds.), *Advances in Social Cognition* (Vol. 2, pp. 1-59). Hillsdale, N.J.: Erlbaum.

Cantor, N., Norem, J.K., Niedenthal, P.M., Langston, C.A., and Brower, A.M. (1987). Life tasks, self-concept ideals, and cognitive strategies in a life transition. *Journal of Personality and Social Psychology*, *53*, 1178-1191.

Carraher, T.N., Carraher, D., and Schliemann, A.D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, *3*, 21-29.

Caruso, D.R., and Mayer, J.D. (1997). A quick scale of empathy. Unpublished manuscript.

Cattell, R.B. (1940). A culture free intelligence test. I. *Journal of Educational Psychology*, *31*, 161-180.

Cattell, R.B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cattell, R.B., and Cattell, H.E.P. (1963). *Test of g: Culture Fair, Scale 3*. Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R.B., and Cattell, H.E.P. (1973). *Measuring intelligence with the Culture Fair Tests*. Champaign, IL: Institute for Personality and Ability Testing.

Ceci, S.J. (1990). *On intelligence. . .more or less: A bio-ecological treatise on cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Ceci, S.J., and Brofenbrenner, U. (1985). Don't forget to take the cupcakes out of the oven: Strategic time-monitoring, prospective memory and context. *Child Development*, *56*, 175-190.

Ceci, S.J., and Liker, J. (1986). Academic and nonacademic intelligence: an experimental separation. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 119-142). New York: Cambridge University Press.

Ceci, S.J., and Liker, J. (1988). Stalking the IQ-expertise relationship: When the critics go fishing. *Journal of Experimental Psychology: General*, *117*, 96-100.

Ceci, S.J. and Roazzi, A. (1994). The effects of context on intelligence: postcards from Brazil. In R.J. Sternberg and R.K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 74-101). New York: Cambridge University Press.

Ceci, S.J., and Ruiz, A. (1991). Cognitive complexity and generality: A case study. In R. Hoffman (Ed.), *The psychology of expertise*. New York: Springer-Verlag.

Chan, D., and Schmitt, N. (1998). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143-159.

Chapin, F.S. (1942). Preliminary standardization of a social impact scale. *American Sociological Review, 7,* 214-225.

Chapin, F.S. (1967). *The Social Insight Test.* Palo Alto, CA: Consulting Psychologists Press.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology, 33,* 1-16.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 467-476.

Charness, N. (1991). Expertise in chess: the balance between knowledge and search. In K.A. Ericsson and J. Smith (Eds.), *Toward a general theory of expertise* (pp. 39-63). New York: Cambridge University Press.

Charness, N., and Bieman-Coplan, S. (1994). The learning prospective: Adulthood. In R.J. Sternberg and C.A. Berg (Eds.), *Cognitive Development* (pp. 301-327). New York, NY: Cambridge University Press.

Charness, N., and Bosman, E.A. (1990). Expertise and aging: Life in the lab. In T.M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 343-385). Amsterdam: Elsevier Science.

Charness, N., and Bosman, E.A. (1995). Compensation through environmental modification. In R.A. Dixon and L. Baeckman, (Eds). *Compensating for psychological deficits and declines: Managing losses and promoting gains.* (pp. 147-168). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Charness, N., Krampe, R., and Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K.A. Ericsson (Ed.), *The road to excellence* (pp. 51-80). Hillsdale, NJ: Erlbaum.

Chase, W.G., and Ericsson, K.A. (1982). Skill and working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1-58). New York: Academic Press.

Chase, W.G., and Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Chen, M.J. (1994). Chinese and Australian concepts of intelligence. *Psychological Developmental Sociology, 6,* 101-117.

Chen, S.A., and Michael, W.B. (1993). First-order and higher-order factors of creative social intelligence within Guilford's structure-of-intellect model: a reanalysis of a Guilford data base. *Educational and Psychological Measurement, 53,* 619-641.

Chi, M.T.H., and Ceci, S.J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. In H.W. Reese (Ed.), *Advances in child development and behavior, Vol. 20* (pp. 91-142). Orlando, FL: Academic Press.

Chi, M.T.H., Glaser, R., and Farr, M.J. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Cohen, M. (1996). Preschoolers' practical thinking and problem solving: the acquisition of an optimal solution strategy. *Cognitive Development, 11,* 357-373.

Cohen, J., and Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Colonia-Willner, R. (1998). Practical intelligence at work: Relationship between aging and cognitive efficiency among managers in a bank environment. *Psychology and Aging, 13,* 45-57.

Cornelius, S.W. (1984). Classic pattern of cognitive aging: Test familiarity, difficulty, and performance. *Journal of Gerontology, 39,* 201-206.

Cornelius, S.W., and Caspi, A. (1987). Everyday problem solving in adulthood and old age. *Psychology and Aging, 2,* 144-153.

Cosier, R.A., and Aplin, J.C. (1982). Intuition and decision making: some empirical evidence. *Psychological Reports, 51,* 275-281.

Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

Csikszentmihalyi, M. (1996). *Creativity*. New York: Harper Collins.

Das, J.P., and Naglieri, J.A. (1997). *Das-Naglieri cognitive assessment system.* Chicago, Illinois: Riverside Publishing Company.

Dasen, P. (1984). The cross-cultural study of cognition: Piaget and the Baoule. *International Journal of Psychology, 19,* 407-434.

Davidson, J.E., and Sternberg, R.J. (1984). The role of insight in cognitive giftedness. *Gifted Child Quarterly, 28,* 58-64.

Davidson, J.E., and Sternberg, R.J. (1998). Smart problem solving: How metacognition helps. In D.J. Hacker, A.C. Graesser, and J. Dunlosky (Eds.), *Metacognition in educational theory and practice.* Mahwah, NJ: Lawrence Erlbaum Associates.

Davies, J., and Easterby-Smith, M. (1984). Learning and developing from managerial work experiences. *Journal of Management Studies, 2,* 169-183.

Davies, M., Stankov, L., and Roberts, R.D. (1998). What remains of emotional intelligence? *Journal of Personality and Social Psychology: Attitudes and Social Cognition, 75,* 989-1015.

De Groot, A. (1978). *Thought and choice in chess.* The Hague: Mouton. (Original work published 1946.)

Denney, N.W. (1979). Problem solving in late life: Intervention research. In P.B. Baltes and O.G. Brim (Eds.), *Life-span development and behavior* (Vol. 2., pp. 37-66). New York: Academic Press.

Denney, N.W. (1982). Aging and cognitive changes. In B.B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 807-827). Englewood Cliffs, NJ: Prentice-Hall.

Denney, N.W. (1989). Everyday problem solving: Methodological issues, research findings, and a model. In I.W. Poon, D.C. Rubin, and B.A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 330-351). New York, NY: Cambridge University Press.

Denney, N.W., and Palmer, A.M. (1981). Adult age differences on traditional and practical problem-solving measures. *Journal of Gerontology, 36,* 323-328.

Denney, N.W., and Pearce, K.A. (1989). A developmental study of practical problem sovling in adults. *Psychology and Aging, 4,* 438-442.

Denney, N.W., Pearce, K.A., and Palmer, A.M. (1982). A developmental study of adults' performance on traditional and practical problem-solving tasks. *Experimental Aging Research, 8,* 115-118.

Dixon, R.A. (1994). Contextual approaches to adult cognitive development. In R.J. Sternberg and C.A. Berg (Eds.), *Cognitive Development* (pp. 350-380). New York, NY: Cambridge University Press.

Dixon, R.A., and Baltes, P.B. (1986). Toward life-span research on the functions and pragmatics of intelligence. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 203-235). New York: Cambridge University Press.

Dodge, K.A., Pettit, G.S., McClaskey, C.L., and Brown, M.M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*, 1-85.

Dörner, D., and Kreuzig, H. (1983). Problemlosefahigkeit und intelligenz. *Psychologische Rundschaus, 34,* 185-192.

Dörner, D., Kreuzig, H., Reither, F., and Staudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheir und Komplexitat.* Bern: Huber.

Dreary, I., and Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51,* 599-608.

Eddy, A.S. (1988). *The relationship between the Tacit Knowledge Inventory for Managers and the Armed Services Vocational Aptitude Battery*. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Egan, D.E., and Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7,* 149-158.

Engle, R.W., and Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology, 91,* 673-679.

Ericsson, K.A. (1996). The acquisition of expert performance. In K.A. Ericsson (Ed.), *The road to excellence* (pp. 1-50). Hillsdale, NJ: Erlbaum.

Ericsson, K.A., Krampe, R.T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Ericsson, K.A., and Smith, J. (1991). Empirical study of expertise: Prospects and limits. In K.A. Ericsson and J.A. Smith (Eds.), *Toward a general theory of expertise* (pp. 1-38). New York: Cambridge University Press.

Farah, M.J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307-317.

Finke, R.A., Ward, T.B., and Smith, S.M. (1992). *Creative cognition: theory, research, and applications.* Boston, MA: MIT Press.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

Flavell, J.H. (1970). Developmental studies of mediated memory. In H.W.Reese and L.P. Lipsitt (Eds.), *Advances in child development and child behavior* (Vol 5., pp. 181-211). New York: Academic Press.

Fleeson, W., and Cantor, N. (1995). Goal relevance and the affective experience of daily life: Ruling out situation explanation. *Motivation and Emotion, 19,* 25-57.

Folkman, S., Lazarus, R.S., Pimley, S., and Novacek, J. (1987). Age differences in stress and coping processes. *Psychology and Aging, 2,* 171-184.

Ford, D.H. (1987). *Humans as self-constructing living systems: A developmental perspective on behavior and personality*. Hillsdale, NJ: Erlbaum.

Ford, D.H. (1994). *Humans as self-constructing living systems: A developmental perspective on behavior and personality* (2^{nd} ed.). State College, PA: Ideals.

Ford, M.E. (1982). Social intelligence and social competence in adolescence. *Developmental Psychology, 18,* 323-340.

Ford, M.E. (1986). For all practical purposes: Criteria for defining and evaluating practical intelligence. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 183-200). New York, NY: Cambridge University Press.

Ford, M.E., and Ford, D.H. (1987). *Humans as self-constructing living systems: Putting the framework to work.* Hillsdale, NJ: Erlbaum.

Ford, M.E., and Maher, M.A. (1998). Self-awareness and social intelligence. In M.D. Ferrari and R.J. Sternberg (Eds.), *Self-awareness: Its nature and development* (pp 191-218). New York: The Guilford Press.

Ford, M.E., and Tisak, M.S. (1983). A further search for social intelligence. *Journal of Educational Psychology, 75,* 196-206.

Frederiksen, N. (1966). Validation of a simulation technique. *Organizational Behavior and Human Performance, 1,* 87-109.

Frederiksen, N. (1986). Toward a broader conception of human intelligence. *American Psychology, 41,* 445-452.

Frederiksen, N., Carlson, S., and Ward, W.C. (1984). The place of social intelligence in a taxonomy of cognitive abilities. *Intelligence, 8,* 315-337.

Frederiksen, N., Jensen, O., and Beaton, A.E. (1972). *Prediction of organizational behavior.* New York: Pergamon Press.

Frederiksen, N., Saunders, D.R., and Wand, B. (1957). The in-basket test. *Psychological Monographs, 71.*

Freeman, N.H., Lewis, C., and Doherty, M.J. (1991). Preschoolers' grasp of a desire for knowledge in false-belief prediction: Practical cognition and verbal report. *British Journal of Developmental Psychology, 9,* 139-157.

French, J., and Raven, B.H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies of social power*. Ann Arbor, MI: Institute for Social Research.

Galton, F. (1883). *Inquiry into human faculty and its development.* London: Macmillan.

Gardiner, J.M. (1988). Functional aspects of recollective experience. *Memory and Cognition, 16,* 309-313.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic.

Gardner, H. (1993). Seven creators of the modern era. In J. Brockman (Ed.), *Creativity* (pp. 28-47). New York: Simon and Schuster.

Gardner, H. (1998). Are there additional intelligences? The case for naturalist, spiritual, and existential intelligences. In J. Kane (Ed.), *Education, information, and transformation*. Englewood Cliffs, NJ: Prentice-Hall.

Gardner, H. (1999). Who owns intelligence? *The Atlantic Monthly, 283,* 67-76.

Gardner, H., Krechevsky, M., Sternberg, R.J., and Okagaki, L. (1994). Intelligence in context: Enhancing students' practical intelligence for school. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 105-127). Cambridge, MA: Bradford Books.

Gazzaniga, M.S. (1985). *The social brain: Discovering the networks of the mind.* New York: Basic Books.

Getzels, J., and Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problemfinding in art.* New York: WileyInterscience.

Gill, R., and Keats, D.M. (1980). Elements of cognitive competence: Judgments by Australian and Malay university students. *Journal of Cross-Cultural Psychology, 11,* 233-243.

Goleman, D. (1995). *Emotional intelligence.* New York: Bantam Books.

Goodnow, J.J. (1986). Some lifelong everyday forms of intelligent behavior: Organizing and reorganizing. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence* (pp. 31-50). New York: Cambridge University Press.

Gough, H.G. (1966). Appraisal of social maturity by means of the CPI. *Journal of Abnormal Psychology, 71,* 189-195.

Greenspan, S. (1981). Defining childhood social competence: A proposed working model. In B.K. Keogh (Ed.), *Advances in special education* (Vol. 3, pp. 1-39). Greenwich, CT: JAI Press.

Greenspan, S., and Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D.P. Flanagan and J.L. Genshaft (Eds.), *Contemporary cognitive assessment: Theories, tests, and issues* (pp. 131-150). New York: Guilford Press.

Greenspan, S., and Granfield, J.M. (1992). Reconsidering the construct of mental retardation: Implications of a model of social competence. *American Journal on Mental Retardation, 96,* 442-453.

Grigorenko, E.L., Geissler, P.W., Prince, R., Okatcha, F., Nokes, C., Kenny, D.A., Bundy, D.A., and Sternberg, R.J. (in press). The organization of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development*.

Grigorenko, E.L., and Sternberg, R.J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75-11.

Grigorenko, E.L., and Sternberg, R.J. (in press). Analytical, creative, and practical intelligence as predictors of self-reported adaptive functioning: A case study in Russia. *Intelligence.*

Groen, G.J., and Patel, V.L. (1988). The relationship between comprehension and reasoning in medical expertise. In M.T. H. Chi, R. Glaser, and M. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Guilford, J.P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guilford, J.P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89,* 48-59.

Gustafsson, J.E. (1988). Hierarchical models of individual differences in cognitive abilities. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35-71). Hillsdale, NJ: Erlbaum.

Haier, R.J., Nuechterlein, K.H., Hazlett, E., Wu, J.C., Pack, J., Browning, H.L., and Buchsbaum, M.S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12,* 199-217.

Haier, R.J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M.S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence, 16,* 415-426.

Halberstadt, A.G., and Hall, J.A. (1980). Who's getting the message?: Children's nonverbal abilities and their evaluation by teachers. *Developmental Psychology, 16,* 564-573.

Harter, S. (1983). Developmental prospectives on the self-system. In P.H. Mussen (Ed.), *Handbook of child psychology, Vol. IV.* New York, NY: Wiley.

Hartigan, J.A. (1975). *Clustering algorithms.* New York: Wiley.

Hartley, A.A. (1989). The cognitive etiology of problem solving. In L.W. Poon, D.C. Rubin, and B.A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 300-329). New York, NY: Cambridge University Press.

Havinghurst, R. (1972). *Developmental tasks and education*. New York: Van Nostrand.

Hayes-Roth, F., Waterman, D.A., and Lenat, D.B. (1983). An overview of an expert system. In F. Hayes-Roth, D.A. Watterman, and D.B. Lenat (Eds.), *Building expert systems*. Reading, MA: Addison-Wesley.

Hayslip, B., Jr. (1989a). Alternative mechanisms for improvements in fluid ability performance among older adults. *Psychology and Aging, 4,* 122-124.

Hayslip, B., Jr. (1989b). Fluid ability training with aged people: A past with a future? *Educational Gerontology, 15,* 573-595.

Hebb, D.O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Heckhausen, J., and Schulz, R. (1995). A life-span theory of control. *Psychological Review, 102,* 284-304.

Hedlund, J., Forsythe, G.B., Horvath, J.A., Williams, W.M., Snook, S., Dennis, M., and Sternberg, R.J. (1999). Practical intelligence: The role of tacit knowledge in understanding leadership. Manuscript submitted for publication.

Hedlund, J., Horvath, J.A., Forsythe, G.B., Snook, S., Williams, W.M., Bullis, R.C., Dennis, M., and Sternberg, R.J. (1998). *Tacit Knowledge in Military Leadership: Evidence of Construct Validity* (Tech. Rep. 1080). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Hendricks, M., Guilford, J.P., and Hoepfner, R. (1969) Measuring creative social intelligence. *Reports from the Psychological Laboratory, University of Southern California,* No.42.

Hendrickson, A.E., and Hendrickson, D.E. (1980). The biological basis for individual differences in intelligence. *Personality and Individual Differences, I,* 3-33.

Hofland, B.F., Willis, S.L., and Baltes, P.B. (1981). Fluid intelligent performance in the elderly: Intraindividual variability and conditions of assessment. *Journal of Educational Psychology, 73,* 573-586.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33,* 307-316.

Horn, J.L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F.I.M. Craik and A. Trehum (Eds.), *Aging and cognitive processes* (pp. 237-278). New York: Plenum.

Horn, J.L. (1994). Theory of fluid and crystallized intelligence. In R.J. Sternberg (Ed.), *The encyclopedia of human intelligence* (Vol. 1, pp. 443-451). New York: Macmillan.

Horn, J.L., and Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253-270.

House, R.J. (1971). A path-goal theory of leadership effectiveness. *Administrative Science Quarterly, 16,* 321-339.

Howell, W.C., and Cooke, N.J. (1989). Training the human information processor: A review of cognitive models. In I.L. Goldstein (Ed.), *Training and development in organizations* (pp. 121-182). San Francisco: Jossey-Bass.

Hoyer, W.J., Labouview, G.V., and Baltes, P.B. (1973). Modification of response and speed deficits and cognitive performance in the elderly. *Human Development, 16,* 233-242.

Hunt, T. (1928). The measurement of social intelligence. *Journal of Applied Psychology, 12,* 317-334.

Hunt, E.B. (1978). Mechanics of verbal ability. *Psychological Review, 85,* 109-130.

Hunt, E., Frost, N., and Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 87-122). New York: Academic Press.

Hunt, E.B., Lunneborg, C., and Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology, 7,* 194-227.

Hunter, J.E., and Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-98.

Ilgen, D.R., Barnes-Farrell, J.L., and McKellin, D.B. (1993). Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes, 54,* 321-368.

Jacoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21-38.

Jensen, A.R. (1980*). Bias in mental testing.* New York: Free Press.

Jensen, A.R. (1982). Reaction time and psychometric g. In H.J. Eysenck (Ed.), *A model for intelligence.* Heidelberg: Springer-Verlag.

Jensen, A.R. (1993). Test validity: *g* versus "tacit knowledge," *Current Directions in Psychological Science, 1,* 9-10.

Jones, K. and Day, J.D. (1997). Discrimination of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology, 89,* 486-497.

Kanfer, R., and Ackerman, P.L. (1989). Motivation and cognitive skills: An integrative aptitude—treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74,* 657-690.

Karmiloff-Smith, A. (1988). The child is a theorist not an inductivist. *Mind and Language, 3,* 183-196.

Kaufman, A.S., and Kaufman, N.L. (1993). *Kaufman Adolescent and Adult Intelligence Test.* Circle Pines, MN: American Guidance Service.

Keating, D.K. (1978). A search for social intelligence. *Journal of Educational Psychology, 70,* 218-233.

Kepner, C.H., and Tregoe, B.B. (1965). *The rational manager: A systematic approach to problem solving and decision making.* New York: McGraw-Hill.

Kerr, S., and Jermier, J.M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance, 22,* 375-403.

Keys, B., and Wolfe, J. (1990). The role of management games and simulations in education and research. *Journal of Management, 16,* 307-336.

Kihlstrom, J.F., and Cantor, N. (in press). Social intelligence. In R.J. Sternberg (Ed.), *Handbook of intelligence,* (2nd ed.). Cambridge, UK: Cambridge University Press.

Klaczynski, P.A., Laipple, J.S., and Jurden, F.H. (1992). Educational context differences in practical problem solving during adolescence. *Merrill-Palmer Quarterly, 38,* 417-438.

Kohlberg, L. (1963). The development of children's orientations toward a moral order: I. Sequence in the development of moral thought. *Vita Humana, 6,* 11-33.

Kohn, M.L., and Schooler, C. (Eds.) (1983). *Work and personality*. Norwood, NJ: Ablex.

Kosmitzki, C., and John, O.P. (1993). The implicit use of explicit conceptions of social intelligence. *Personality and Individual Differences, 15,* 11-23.

Kreitler, S., and Kreitler, H. (1987). Conceptions and processes of planning: The developmental perspective. In S.L. Friedman and E.K. Scholnick (Eds.), *Blueprints for thinking: The role of planning in cognitive development* (pp. 205-272). Cambridge, UK: Cambridge University Press.

Kuhn, D., Pennington, N., and Leadbeater, B. (1983). Adult thinking in developmental perspective. In P.B. Baltes and O.G. Brim (Eds.), *Life-span development and behavior* (Vol. 5). New York: Academic Press.

Laboratory of Comparative Human Cognition (1982). Culture and intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 642-719). New York: Cambridge University Press.

Labouvie-Vief, G. (1982). Dynamic development and mature autonomy. *Human Development, 25,* 161-191.

Labouvie-Vief, G. (1992). A Neo-Piagetian perspective on adult cognitive development. In R.J. Sternberg and C.A. Berg (Eds.), *Cognitive development* (pp. 197-228). New York, NY: Cambridge University Press.

Laipple, J.S. (1992). Problem-solving in young and old adulthood: The role of task interpretation. *Dissertation Abstracts International, 53(1-B),* 582.

Landy, F.J., and Farr, J. (1980). Performance ratings. *Psychological Bulletin, 87,* 72-107.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life.* New York: Cambridge University Press.

Lave, J., Murtaugh, M., and de la Roche, O. (1984). The dialectic of arithmetic in grocery shopping. In B. Rogoff and J. Lace (Eds.), *Everyday cognition: Its development in social context* (pp. 67-94). Cambridge, MA: Harvard University Press.

Lazarus, R.S., and Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer.

Legree, P.J. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence, 21,* 247-266.

Lesgold, A., Rubinson, H. Feltovich, P., Glaser, R., Klopfer, D., and Wang, Y. (1985). *Expertise in a complex skill: Diagnosing X-ray pictures.* Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh Technical Report.

Likert, R. (1967). The human organization: Its management and value. New York: McGraw-Hill.

Loewen, E.R., Shaw, J.R., and Craik, F.I.M. (1990). Age differences in components of metamemory. *Experimental Aging Research, 16,* 43-48.

Manz, C.C., and Sims, H.P. Jr. (1980). Self-management as a substitute for leadership: A social learning theory perspective. *Academy of Management Review, 5,* 361-367.

Marlowe, H.A. (1986). Social intelligence: Evidence for multidimensionality and construct independence. *Journal of Educational Psychology, 78,* 52-58.

Mayer, J.D., Caruso, D., and Salovey, P. (in press). Emotional intelligence meets traditional standards for an intelligence. test. *Intelligence.*

Mayer, J.D., and Greeno, J.G. (1972). Structural differences between learning outcomes produced by different instructional methods. *Journal of Educational Psychology, 63,* 165-173.

Mayer, J.D., and Salovey, P. (1993). The intelligence of emotional intelligence. Intelligence*, 17*, 433-442.

Mayer, J.D., and Salovey, P. (1997). What is emotional intelligence? In P. Salovey and D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3-31). New York: Basic Books.

Mayer, J.D., Salovey, P., and Caruso, D. (in press). Competing models of emotional intelligence. In R.J. Sternberg (Ed)., *Handbook of intelligence* (2nd ed.). New York: Cambridge University Press.

McCall, M.W., and Kaplan, R.E. (1985). *Whatever it takes: Decision makers at work.* Englewood Cliffs, NJ: Prentice-Hall.

McClelland, D.C. (1973). Testing for competence rather than for "intelligence." *American Psychologist, 28,* 1-14.

McClelland, D.C. (1976). *A guide to job competency assessment.* Boston: McBer.

Meacham, J.A. and Emont, N.C. (1989). The interpersonal basis of everyday problem solving. In J.D. Sinnott (Ed.), *Everyday problem solving* (pp. 7-23). New York, NY: Praeger.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50 (9),* 741-750.

Miller, K.A., and Kohn, M.L. (1983). The reciprocal effects on job conditions and the cognitivity of leisure0time activities. In M.L. Kohn and C. Schooler (Eds.), *Work and personality* (pp. 217-241). Norwood, NJ: Ablex.

Minsky, M. (1968). *Semantic information processing.* Cambridge, MA: MIT Press.

Mintzberg, H., Raisinhani, D., and Theoret, A. (1976). The structure of "unstructured" decision processes. *Administration Science Quarterly, 21,* 246-275.

Mischel, W. (1984). Convergences and challenges in the search for consistency. *American Psychologist, 39,* 351-364.

Moss, F.A., and Hunt, T. (1927) Are you socially intelligent? *Scientific American, 137,* 108-110.

Moss, F.A., Hunt, T., Omwake, K.T., and Woodward, L.G. (1949). *Social Intelligence Test, George Washington University Series.* Washington, DC: Center for Psychological Services, George Washington University.

Motowidlo, S.J., Dunnette, M.D., and Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.

Murtaugh, M. (1985). The practice of arithmetic by American grocery shoppers. *Anthropology and Education Quarterly, 16,* 186-192.

Naus, M.J. and Ornstein, P.A. (1983). Development of memory strategies: Analysis, questions and issues. In M.T.M. Chi (Ed.), *Trends in memory development research: Contributions to human development* (Vol. 9, pp. 1-30). Basel, Switzerland: S. Karger.

Neely, A.S., and Backman, L. (1993). Long-term maintenance of gains from memory training in older adults: Two 3 ½-year follow-up studies. *Journal of Gerontology, 48,* 233-237.

Neisser, U. (1976). General, academic, and artificial intelligence. In L. Resnick (Ed.), *Human intelligence: Perspectives on its theory and measurement* (pp. 179-189). Norwood, NJ: Ablex.

Neisser, U. (1979). The concept of intelligence. In R.J. Sternberg and D.K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 179-189). Norwood, NJ: Ablex.

Nelson, M.J., and Lamke, T.A. (1973). *The Henmon-Nelson test of mental ability*. Boston: Houghton Mifflin.

Nettlebeck, T. (1982). Inspection time: An index for intelligence. *Quarterly Journal of Experimental Psychology, 34,* 299-312.

Neugarten, B.L., Moore, J.W., and Lowe, J.C. (1968). Age norms, age constraints, and adult socialization. In B.L. Neugarten (Ed.), *Middle age and aging* (pp. 22-28). Chicago: University of Chicago Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., and Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nuñes, T. (1994). Street intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp 1045-1049). New York: Macmillan.

Nuñes, T., Schliemann, A.D., and Carraher, D.W. (1993). *Street mathematics and school mathematics.* New York: Cambridge University Press.

Nunally, J.C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Okagaki, L., and Sternberg, R.J. (1993). Parental beliefs and children's school performance. *Child Development, 64(1),* 36-56.

Oltman, P.K., Raskin, F., and Witkin, H.A. (1971). *Group Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press.

O'Sullivan, M., and Guilford, J.P. (1976). *Four factor tests of social intelligence (behavioral intelligence): Manual of instructions and interpretations.* Orange, CA: Sheridan Psychological Services.

O'Sullivan, M., Guilford, J.P., and deMille, R. (1965). The measurement of social intelligence. *Reports from the Psychological Laboratory, University of Southern California*, No. 34.

Patel, V.L., and Groen, G.J. (1991). The general and specific nature of medical expertise: a critical look. In K.A. Ericsson and J. Smith (Eds.), *Toward a general theory of expertise* (pp. 93-125). Hillsdale, NJ: Erlbaum.

Perret-Clermont, A.N. (1980). *Social interaction and cognitive development in children.* London: Academic Press.

Piaget, J. (1972). *The psychology of intelligence.* Totowa, NJ: Littlefield Adams.

Plunkett, L.C., and Hale, G.A. (1982). *The proactive manager.* New York: John Wiley and Sons.

Polanyi, M. (1966). *The tacit dimensions.* Garden City, NY: Doubleday.

Polanyi, M. (1976). Tacit knowledge. In M. Marx and F. Goodson (Eds.), *Theories in contemporary psychology* (pp. 330-344). New York: Macmillan.

Pressley, M., Forrest-Presley, D.L., Elliot-Faust, D., and Miller, G. (1985). Children's use of cognitive strategies: How to teach strategies, and what to do if they can't be taught. In M. Pressley and C.J. Brainers (Eds.), *Cognitive learning and memory in children: Progress in cognitive development research* (pp. 1-47). New York: Springer.

Raven, J.C. (1958). *Guide to using the Coloured Progressive Matrices*. London, England: H.K. Lewis and Co.

Raven, J.C., Court, J.H., and Raven, J. (1992). *Manual for Raven's Progressive Matrices and Mill Hill Vocabulary Scales.* Oxford: Oxford Psychologists Press.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 317-327.

Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219-235.

Reber, A.S., and Millward, R.B. (1968). Event observation in probability learning. *Journal of Experimental Psychology, 77,* 317-327.

Ree, M.J. and Earles, J.A. (1993). *g* is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee, *Current Directions in Psychological Science, 1,* 11-12.

Reed, T.E. and Jensen, A.R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16,* 259-272.

Reichert, U., and Dorner, D. (1988). Heurismen beim Umgang mit einem "einfachen" dynamischen System [Heuristics in the control of a "simple" dynamic system]. *Sprache and Kognition, 7,* 12-24.

Reitman, J. (1976). Skilled perception in GO: Deducing memory structures from interresponse times. *Cognitive Psychology, 8,* 336-356.

Rest, J. (1975). Longitudinal study of the Defining Issues Test of moral judgement: A strategy for analyzing developmental change. *Developmental Psychology, 11,* 738-748.

Riggio, R.E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology, 51,* 649-660.

Riggio, R.E. (1989). *Manual for the Social Skills Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Riggio, R.E., Messamer, J., and Throckmorton, B. (1991). Social and academic intelligence: Conceptually distinct but overlapping constructs. *Personality and Individual Differences, 12,* 695-702.

Roazzi, A. (1987). Effects of context on cognitive development. In J.F. Cruz and R.A. Goncalves (Eds.), *Psicologia e Educao: Investigacao e intervencao.* Porto: Associacao dos Piscologos Portugueses.

Rogoff, B. (1982). Integrating context and cognitive development. In M.E. Lamb and A.L. Brown (Eds.), *Advances in development psychology* (Vol. 2, pp. 125-169). Hillsdale, NJ: Erlbaum.

Rogoff, B., Gauvain, M., and Gardner, W. (1987). Children's adjustment of plans to circumstances. In S.L. Friedman, E.K., Scholnick, and R.R. Cocking (Eds.), *Blueprints for thinking* (pp. 303-320). New York, NY: Cambridge University Press.

Rogoff, B. and Lave, J. (Eds.). (1984). *Everyday cognition: its development in social context.* Cambridge, Massachusetts: Harvard University Press.

Rosenthal, R. (Ed.) (1979). *Skill in nonverbal communication: Individual differences.* Cambridge, MA: Oelgeschlager, Gunn, and Hain.

Rosenthal, R., Hall, J.A., DiMatteo, M.R., Rogers, P.L., and Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test.* Baltimore: Johns Hopkins University Press.

Ross, L.D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press.

Salovey, P., and Mayer, J.D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9,* 185-211.

Salovey, P., and Mayer, J.D. (1994). Some final thoughts about personality and intelligence. In R.J. Sternberg, and P. Ruzgis (Eds.), *Personality and intelligence* (pp. 303-318). Cambridge, UK: Cambridge University Press.

Salthouse, T.A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General, 113,* 345-371.

Salthouse, T.A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Salthouse, T.A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26,* 153-166.

Salthouse, T.A., and Somberg, B.L. (1982). Skilled performance: The effects of adult age and experience on elementary processes. *Journal of Experimental Psychology: General, 111,* 176-207.

Sansone, C., and Berg, C.A. (1993). Adapting to the environment across the life span: Different process or different inputs? *International Journal of Behavioral Development, 16,* 215-241.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology, 13,* 501-518.

Schaie, K.W. (1977/1978). Toward a stage theory of adult cognitive development. *International Journal of Aging and Human Development, 8,* 129-138.

Schaie, K.W. (1986). Beyond calendar definitions of age, time, and cohort: The general developmental model revisited. *Developmental Review*, *6*, 252-277.

Schaie, K.W. (1988). Variability in cognitive functioning in the elderly: Implications for societal participation. In A.D. Woodhead, M.A. Bender, and R.C. Leonard (Eds.), *Phenotypic variation in populations: Relevance to risk assessment* (pp. 191-212). New York: Plenum.

Schaie, K.W. (1989). The hazards of cognitive aging. *Gerontologist*, *29*, 484-493.

Schaie, K.W. (1994). The course of adult development. *American Psychologist*, *49*, 304-313.

Schaie, K.W. (1996). *Cognitive development in adulthood: The Seattle Longitudinal Study*. New York, NY: Cambridge University Press.

Schaie, K.W. and Willis, S.L. (1986). Can decline in adult cognitive functioning be reversed? *Developmental Psychology*, *22*, 223-232.

Schank, R.C. (1990). *Tell me a story: A new look at real and artificial memory*. New York: Charles Scribner's Sons.

Schmidt, F.L., and Hunter, J.E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge, *Current Directions in Psychological Science*, *1*, 8-9.

Schmidt, F.L, and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Schooler, C. (1984). Psychological effects of complex environments during the life span: A review and theory. *Intelligence*, *8*, 259-281.

Schooler, C. (in press). The cognitive effects of the demands of the work environment. In R.S. Sternberg and E.L. Grigorenko (Eds.) *Environmental effects on cognitive functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Schon, D.A. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Schutte, N.S., Malouff, J.M., Hall, L.E., Haggerty, D.J., Cooper, J.T., Golden, C.J., and Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, *25*, 167-177.

Scribner, S. (1984). Studying working cognition. In B. Rogoff and J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9-40). Cambridge, MA: Harvard University Press.

Scribner, S. (1986). Thinking in action: Some characteristics of practical thought. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 13-30). New York: Cambridge University Press.

Scribner, S., and Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Serpell, R. (1982). Measures of perception, abilities, and intelligence. In W.W. Hartup (Ed.), *Review of Child Development Research.* Chicago: University of Chicago Press.

Shiffrin, R.M. (1996). Laboratory experimentation on the genesis of expertise. In K.A. Ericsson (Ed.), *The road to excellence* (pp. 337-345).

Shipley, W.C. (1940). A self-administering scale for measuring cognitive impairment and deterioration. *Journal of Psychology, 9,* 371-377.

Shirley, D.A., and Langan-Fox, J. (1996). Intuition: a review of the literature. *Psychological Review, 79,* 563-684.

Siegler, R.S., and Richards, D.D. (1982). The development of intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence.* New York: Cambridge University Press.

Simon, H.A., and Chase, W.G. (1973). Skill in chess. *American Scientist, 61,* 391-403.

Simon, H.A., and Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 8,* 165-190.

Simonton, D.K. (1980). Intuition process and analysis: a predictive and explanatory model. *Genetic Psychology Monographs, 102,* 3-60.

Simonton, K.A. (1996). Creative expertise: A life-span developmental perspective. In K.A. Ericsson (Ed.), *The road to excellence* (pp. 227-253).

Sinnott, J.D. (1989). A model for solution of ill-structured problems: implications for everyday and abstract problem solving. In J.D. Sinnott (Ed.), *Everyday problem solving: Theory and applications.* New York: Praeger.

Smith, J. and Baltes, P.B. (1990). Wisdom-related knowledge: Age/cohort differences in response to life-planning problems. *Developmental Psychology, 26,(3),* 494-505.

Smith, P.C., and Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149-155.

Snow, R.E., and Lohman, D.F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347-376.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30,* 526-537.

Spearman, C.E. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology, 15,* 201-293.

Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition* (2nd ed.). London: Macmillan. (1923 edition reprinted in 1973 by Arno Press, New York).

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Sperry, R.W. (1961). Cerebral organization and behavior. *Science, 133,* 1749-1757.

Sternberg, R.J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology, 73,* 1-16.

Sternberg, R.J. (1983). Components of human intelligence. *Cognition, 15,* 1-48.

Sternberg, R.J. (1985a). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1985b). *Human abilities: An information-processing approach*. San Francisco: Freeman.

Sternberg, R.J. (1986). *Intelligence applied: Understanding and increasing your cognitive abilities*. San Diego: Harcourt Brace Jovanovich.

Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown and M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Sternberg, R.J. (1990a). *Metaphors of mind: Conceptions of the nature of intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1991a). Theorybased testing of cognitive abilities: Rationale for the Triarchic Abilities Test. In H. Rowe (Ed.), *Cognition: Reconceptualization and measurement* (pp. 183-202). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J. (1991b). Triarchic abilities test. In D. Dickinson (Ed.), *Creating the future: Perspectives on educational change* (pp. 76-81). Aston Clinton, Bucks, UK: Accelerated Learning Systems.

Sternberg, R.J. (1993). *Sternberg Triarchic Abilities Test.* Unpublished test.

Sternberg, R.J. (1994a). Cognitive conceptions of expertise. *International Journal of Expert Systems: Research and Applications*, *7* (1), 1-12.

Sternberg, R.J. (1994b). Intelligence. In R.J. Sternberg (Ed.), *Handbook of perception and cognition: Thinking and problem solving* (pp. 263-288). San Diego: Academic Press.

Sternberg, R.J. (1994c). Tacit knowledge and job success. In N. Anderson and P. Herriot (Eds.), *Assessment and selection in organizations* (pp. 27-39). New York: Wiley.

Sternberg, R.J. (1995a). Expertise in complex problem solving: A comparison of alternative conceptions. In P.A. Frensch and J. Funke (Eds.), *Complex problem solving: European perspectives* (pp. 295-321). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J. (1995b). *In search of the human mind.* Orlando: Harcourt Brace College Publishers.

Sternberg, R.J. (1995c). Theory and measurement of tacit knowledge as part of practical intelligence. *Zeitschrift fur Psychologie, 203,* 319-333.

Sternberg, R.J. (1996). Costs of expertise. In K.A. Ericsson (Ed.), *The road to excellence.* Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J. (1997a). *Successful intelligence.* New York: Plume.

Sternberg, R.J. (1997b). Tacit knowledge and job success. In N. Anderson and P. Herriot (Eds), *International handbook of selection and assessment* (pp. 201-213). New York: Wiley.

Sternberg, R.J. (1998a). Abilities as forms of developing expertise. *Educational Researcher, 27,* 11-20.

Sternberg, R.J. (1998c). Successful intelligence: An expanded approach to understanding intelligence. In K. Pribram (Ed.), *Brain and Values: Is a biological science of values possible* (pp.1-21). Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J. (1999). A propulsion model of types of creative contributions. *Review of General Psychology, 3,* 83-100.

Sternberg, R.J. (in press-a). Intelligence as developing expertise. *Contemporary Educational Psychology*.

Sternberg, R.J. (in press-b). The theory of successful intelligence. *Review of General Psychology*.

Sternberg, R.J., and Ben-Zeev, T. (in press). *Complex cognition: The psychology of human thought*. New York: Oxford University Press.

Sternberg, R.J., and Berg, C.A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R.J. Sternberg and D.K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 155-162). Norwood, NJ: Ablex.

Sternberg, R.J., and Berg, C.A. (1992). Intellectual *development*. New York: Cambridge University Press.

Sternberg, R.J., and Clinkenbeard, P.R. (1995). A triarchic model of identifying, teaching, and assessing gifted children. *Roeper Review, 17 (4),* 255-260.

Sternberg, R.J., Conway, B.E., Ketron, J.L., and Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41,* 37-55.

Sternberg, R.J., and Detterman D.K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R.J., Ferrari, M., Clinkenbeard, P.R., and Grigorenko, E.L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly*, *40*(3), 129-137.

Sternberg, R.J., Forsythe, G.B., Hedlund, J., Horvath, J.A., Wagner, R.K., Williams, W.M., Snook, S., and Grigorenko, E.L. (in press). *Practical intelligence in everyday life.* New York: Cambridge University Press.

Sternberg, R.J., and Gardner, M.K. (1982). A componential interpretation of the general factor in human intelligence. In H.J. Eysenck (Ed.), *A model for intelligence* (pp. 231-254). Berlin: SpringerVerlag.

Sternberg, R.J., and Grigorenko, E.L. (1997a). The cognitive costs of physical and mental ill health: Applying the psychology of the developed world to the problems of the developing world. *Eye on Psi Chi, 2,* 20-27.

Sternberg, R.J., and Grigorenko, E.L. (Eds.). (1997b). *Intelligence, heredity and environment.* New York: Cambridge University Press.

Sternberg, R.J., Grigorenko, E.L., and Ferrari, M. (in press). Developing expertise: The respective roles of abilities and deliberate practice. In M. Ferrari (Ed.), *Expertise and its development*. New York: Guilford.

Sternberg, R.J., Grigorenko, E.L., Ferrari, M. and Clinkenbeard, P. (1999). A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment*,*15*(1), 1-11.

Sternberg, R.J., Grigorenko, E.L., Ngorosho, D., Tantubuye, E., Mbise, A., Nokes, C., and Bundy, D.A. (1999). Hidden cognitive potential in rural Tanzanian school children. Unpublished manuscript.

Sternberg, R.J., and Horvath, J.A. (Eds.). (1999). *Tacit knowledge in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R.J., and Kaufman, J.C. (1996). Innovation and intelligence testing: The curious case the dog that didn't bark. *European Journal of Psychological Assessment, 12*, 175-182.

Sternberg, R.J., and Kaufman, J.C. (1998). Human abilities. *Annual Review of Psychology*, *49*, 479-502.

Sternberg, R.J., and Lubart, T.I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity.* New York: Free Press.

Sternberg, R.J., Nokes, K., Geissler, P.W., Prince, R., Okatcha, F., Bundy, D., and Grigorenko, E.L. (in press). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*.

Sternberg, R. J., Okagaki, L., and Jackson, A. (1990). Practical intelligence for success in school. *Educational Leadership, 48,* 35-39.

Sternberg, R.J., and Powell, J.S. (1983). Comprehending verbal comprehension. *American Psychologist, 38,* 878-893.

Sternberg, R.J., and Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. *Social Cognition, 3,* 168-192.

Sternberg, R.J., and Wagner, R.K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world.* New York: Cambridge University Press.

Sternberg, R.J., and Wagner, R.K. (1993). The *g*eocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2,* 1-5.

Sternberg, R.J., and Wagner, R.K. (Eds.). (1994). *Mind in context.* New York: Cambridge University Press.

Sternberg, R.J., Wagner, R.K., and Okagaki, L. (1993). Practical intelligence: The nature and role of tacit knowledge in work and at school. In H. Reese and J. Puckett (Eds.), *Advances in lifespan development* (pp. 205-227). Hillsdale, NJ: Erlbaum.

Sternberg, R.J., Wagner, R.K., Williams, W.M., and Horvath, J.A. (1995). Testing common sense. *American Psychologist, 50,* 912-927.

Strang, R. (1930). Measures of social intelligence. *American Journal of Sociology, 36,* 263-269.

Stricker, L.J., and Rock, D.A. (1990). Interpersonal competence, social intelligence, and general ability. *Personality and Individual Differences, 11,* 833-839.

Strough, J., Berg, C., and Sansone, C. (1996). Goals for solving everyday problems across the life span: age and gender differences in the salience of interpersonal concerns. *Developmental Psychology, 32,* 1106-1115.

Super, C.M., and Harness, S. (1982). The infants' niche in rural Kenya and metropolitan America. In L.L Adler (Ed.), *Cross-Cultural Research at Issue.* New York: Academic Press.

Terman, L.M. (1950). *Concept Mastery Test.* New York: The Psychological Corporation.

Thorndike, E.L. (1920). Intelligence and its use. *Harper's Magazine, 140,* 227-235.

Thorndike, R.L., and Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin, 34,* 275-285.

Thornton, G.C., and Byham, W.C. (1982). *Assessment centers and managerial performance.* New York: Academic Press.

Thornton, G.C., and Cleveland, J.N. (1990). Developing managerial talent through simulation. *American Psychologist, 45,* 190-199.

Tornow, W.W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management, 32,* 221-229.

Vallacher, R.R., and Wegner, D.M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94,* 3-15.

Valsiner, J., and Leung, M.C. (1994). From intelligence to knowledge construction: a sociogenetic process approach. In R.J. Sternberg and R.K. Wagner (Eds.), *Mind in context* (pp. 202-217). New York: Cambridge University Press.

Vernon, P.E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology, 4,* 42-57.

Voss, J.F., Perkins, D.N., and Segal, J.W. (Eds.). (1991). *Informal reasoning and education.* Hillsdale, NJ: Erlbaum.

Vygotsky, L. (1978). *Mind in society: The development of higher order processes.* Cambridge, MA: Harvard University Press.

Wagner, D.A. (1978). Memories of Morocco: The influence of age, schooling and environment on memory. *Cognitive Psychology, 10,* 1-28.

Wagner, R.K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology, 52,* 1236-1247.

Wagner, R.K. (1991). Managerial problem-solving. In R.J. Sternberg and P. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 159-183). Hillsdale, NJ: Erlbaum.

Wagner, R.K. (1997). Intelligence, training, and employment. *American Psychologist, 52,* 1059-1069.

Wagner, R.K. (in press). Practical intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence*. New York, NY: Cambridge University Press.

Wagner, R.K., Rashotte, C.A., and Sternberg, R.J. (1994). *Tacit Knowledge in sales: Rules of thumb for selling anything to anyone.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Wagner, R.K., and Sternberg, R.J. (1985). Practical intelligence in realworld pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49,* 436-458.

Wagner, R.K., and Sternberg, R.J. (1986). Tacit knowledge and intelligence in the everyday world. In R.J. Sternberg and R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 51-83). New York: Cambridge University Press.

Wagner, R.K., and Sternberg, R.J. (1989). *Tacit Knowledge Inventory for Sales: Written*. Unpublished test.

Wagner, R.K., and Sternberg, R.J. (1990). Street smarts. In K.E. Clark and M.B. Clark (Eds.), *Measures of leadership* (pp.493-504). West Orange, NJ: Leadership Library of America.

Wagner, R.K., and Sternberg, R.J. (1991). *Tacit Knowledge Inventory for Managers.* San Antonio, TX: The Psychological Corporation.

Wagner, R.K., Sujan, H., Sujan, M., Rashotte, C.A. and Sternberg, R.J. (1999). Tacit knowledge in sales. In R.J. Sternberg and J.A. Horvath (Eds.), *Tacit knowledge in professional practice* (pp. 155-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams and Wilkins.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Wedeck, J. (1947). The relationship between personality and psychological ability. *British Journal of Psychology*, 36, 133-151.

Wertsch, J.V. (1985). *Vygotsky and the social formation of mind.* Cambridge, MA: Harvard University Press.

Wertsch, J., and Kanner, B.G. (1994). A sociocultural approach to cognitive development. In R.J. Sternberg and C.A. Berg (Eds.), *Cognitive development* (pp. 328-349). New York, NY: Cambridge University Press.

Wigdor, A.K. and Garner, W.R. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies.* Washington, DC: National Academy Press.

Williams, S.A., Denney, N.W., and Schadler, M. (1983). Elderly adults' perception of their own cognitive development during the adult years. *International Journal of Aging and Human Development, 16,* 147-158.

Williams, W.M., Blythe, T., White, N., Li, J., Sternberg, R.J., and Gardner, H.I. (1996). *Practical intelligence for school: A handbook for teachers of grades 5-8.* New York: HarperCollins.

Willis, S.L. (1987). Cognitive training and everyday competence. In K.W. Schaie (Ed.), *Annual Review of Gerontology and Geriatrics* (Vol. 7, pp. 159-188). New York, NY: Springer.

Willis, S.L. (1989). Improvement with cognitive training: Which dogs learn what tricks? In L.W. Poon, D.C. Rubin, and B.A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 300-329). New York, NY: Cambridge University Press.

Willis, S.L. (1990). Contributions of cognitive training research to late life potential. In M. Perlmutter (Ed.), *Late life potential* (pp. 25-42). Washington, DC: Gerontological Society of America.

Willis, S.L., Blieszner, R., and Baltes, P.B. (1981). Training research in aging: Modification of performance on the fluid ability of figural relations. *Journal of Educational Psychology, 73,* 41-50.

Willis, S.L., and Nesselroade, C.S. (1990). Long-term effects of fluid ability training in old-old age. *Developmental Psychology, 26,* 905-910.

Willis, S.L., and Schaie, K.W. (1986). Practical intelligence in later adulthood. In R.J. Sternberg and R. Wagner (Eds.), *Practical intelligence* (pp. 236-270). New York: Cambridge University Press.

Willis, S.L. and Schaie, K.W. (1994). Cognitive training in the normal elderly. In F. Boller (Ed.), *Cerebral plasticity and cognitive stimulation.* New York: Springer-Verlag.

Wong, C.T., Day, J.D., Maxwell, S.E., and Meara, N.M. (1995). A multitrait-multimethod study of academic and social intelligence in college students. *Journal of Educational Psychology, 87,* 117-133.

Worden, P.E., and Sherman-Brown, S. (1983). A word-frequency cohort effect in young versus elderly adults-memory for words. *Developmental Psychology, 19,* 521-530.

Yang, S., and Sternberg, R.J. (1997). Conceptions of intelligence in ancient Chinese philosophy. *Journal of Theoretical and Philosophical Psychology, 17(2),* 101-119.

Yerkes, R.M. (1921). *The measurement and appraisal of adult intelligence (4th ed.).* Baltimore, MD: Williams and Wilkins.

Yussen, S.R., and Kane, P. (1985). Children's concept of intelligence. In S.R. Yussen (Ed.), *The growth of reflection in children.* New York: Academic Press.

Chapter 9

The ICT Literacy Framework

Irwin Kirsch *Educational Testing Service Princeton, NJ*
Marilou Lennon *Educational Testing Service Princeton, NJ*

Materials selected from "*Digital Transformation: A Framework for ICT Literacy.* A report of the International ICT Literacy Panel, Kirsch, Irwin and Lennon, Marilou, Educational Testing Service, May 2002." Reprinted by permission of Educational Testing Service, the copyright owner.

Table of contents

1. Introduction

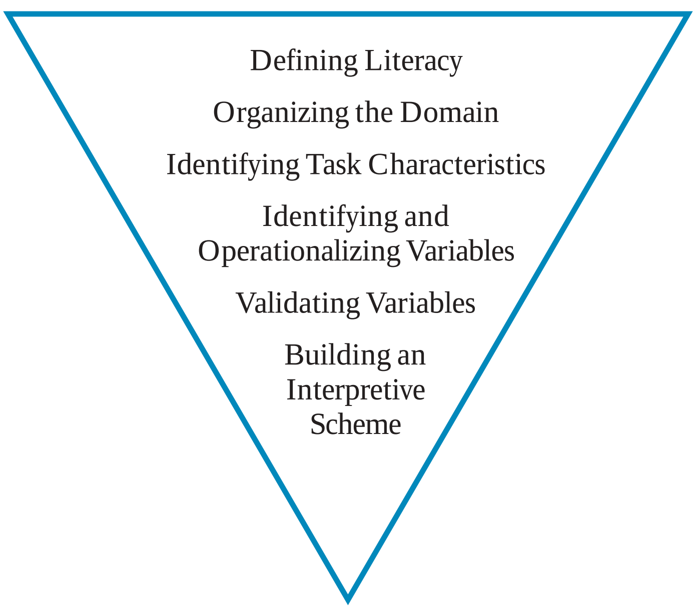
What does it mean to be a literate member of society? The growing acceptance of lifelong learning has expanded the views and demands of literacy. Literacy is no longer seen as a condition that one either has or is lacking. Rather, it is seen as a continuum of knowledge, skills, and strategies that individuals acquire over the course of their lives in various contexts and through interactions with their peers and with the larger communities in which they participate. As historians remind us, literacy in its earliest form consisted of little more than being able to sign one's name on a legal document. It was not until later that fluent oral reading became important and not until the 20^{th} century that reading to gain information was given primary emphasis. As we move into the 21^{st} century, our conception of literacy is evolving once again. The prevalence of technology in the everyday lives of the world's citizens has grown at a rate that many would have found hard to imagine 25 or even 10 years ago. Policy makers, business leaders, and educators have come to expand their notion of a literate populace to include the skills and abilities that will enable citizens to function in an increasingly technological world.

2. Developing a framework

The task of the International ICT Literacy Panel was to develop a framework that would define ICT literacy and provide the foundation for the design and conduct of large-scale assessments and diagnostic tests. While the chief benefit of developing a framework for ICT literacy is improved measurement, a number of other potential benefits are also seen as important. Namely,

- A framework provides a common language and a vehicle for discussing the definition and assumptions surrounding the domain.
- Such a discussion provides a mechanism for building consensus around the framework and measurement goals that grow from it.
- We construct a better understanding of what is being measured through the process of developing the framework and linking it to evidence collected from assessment tasks.
- This understanding and its connection to what we say about learners provides an important link between public policy, assessment, and research which furthers the utility of the data that are collected.

To accomplish this task, the panel chose to adopt the process used to develop frameworks for the International Adult Literacy Survey (OECD and Statistics Canada 1995; OECD and Development and Human Resources Development Canada 1997; OECD and Statistics Canada 2000) and for the Reading Literacy Survey conducted as part of PISA, the Programme for International Student Assessment (OECD, 1999). This process consists of six steps, shown in the following diagram and explained more fully below (Kirsch 2001).



1. The first step is to develop a working definition of the domain including the assumptions underlying it. Before the definition is developed, the domain and the skills and abilities it encompasses are wide open. It is the definition that sets the boundaries for what will be measured and what will not.

2. Once the definition is developed, it is important to think about the kinds of tasks that represent the skills and abilities included under that definition. Those tasks must then be categorized, or organized, to inform test design and result in meaningful score reporting. Step 2 allows one to move beyond a laundry list of tasks or skills to a coherent representation of the domain that will permit policy makers and others to summarize and report information in more useful ways.
3. Step 3 involves identifying a set of key characteristics that will be used in constructing tasks for the assessment. This may include characteristics of the stimulus materials to be used as well as characteristics of the tasks presented to examinees.
4. In step 4, the variables associated with each task characteristic are specified.
5. In step 5, research is conducted to show which variables account for large percentages of the variance in the distribution of tasks and thereby contribute most towards understanding task difficulty and predicting performance.
6. Finally in step 6, an interpretative scheme is built that uses the validated variables to explain task difficulty and examinee performance. The work of this panel involved the first two steps: defining ICT literacy and organizing the domain.

3. Defining ICT literacy

The International ICT Literacy Panel was comprised of educators, technology experts, scholars and industry and labor representatives from Australia, Brazil, Canada, France, and the United States. Our deliberations resulted in the following definition:

> ICT literacy is using digital technology, communications tools, and/or networks to access, manage, integrate, evaluate and create information in order to function in a knowledge society.

This definition carries several assumptions made by the panel and therefore it is important to consider each part of the definition in turn.

"ICT..."

Information Technology (IT) has been used for many years, particularly in the United States, and refers to the electronic display, processing, and storage of information, but not necessarily the transmission of the information. The term carries strong historical associations with enterprise data processing and centralized computer services.

However, Information and Communication Technology (ICT) represents the set of activities and technologies that fall into the union of IT and communication technologies. Global industry, international media, and academics increasingly now use ICT to describe this union. The real benefit of adding "communication" doesn't derive from including specific technologies, such as routers or servers, but from the dynamism implicit in interconnected social, economic, and information networks. ICT is characterized by unprecedented global flows in information, products, people, capital, and ideas. These flows are enabled by ICT: their sheer scale and pace would not be possible without the ability to connect vast networks of individuals across geographic boundaries at negligible marginal cost.

"...literacy is..."

The panel selected the term literacy over other terms such as competency, ability, or fluency that have been used in earlier frameworks (Committee on Information Technology Literacy, 1999). To some "literacy" connotes functional literacy and implies basic or fundamental skills. To the panel, the term literacy implies a universal need, a condition that must be met to enable full and equitable economic and social participation. We view literacy as a tool that may be applied to simple or more complicated contexts — like a hammer that can be used to build a shelf, or a house. In its broadest sense, literacy is a dynamic tool that allows individuals to continuously learn and grow.

The increasing role of technology in our lives requires us to expand our notion of literacy. It is obvious that to function fully and effectively in society, individuals must be literate in terms of traditional domains such as reading and numeracy. But today it is becoming increasingly clear that ICT literacy joins the ranks of essential and fundamental requirements. Perhaps as important is the panel's belief that those who fail to acquire this new kind of literacy, like the more traditional literacy skills, will find themselves falling further behind as economies and societies grow and change over the years ahead.

"...using digital technology, communications tools, and/or networks..."

The description of digital technology, communication tools, and/or networks reflects the same thinking that stimulated the panel's use of information and communication technology (ICT) versus information technology (IT). Digital technology reflects hardware and software products, communication tools reflect those products and services used to transmit information, and networks themselves are the pathways for this transmission. The words are meant to be as inclusive as possible to reflect the breadth of hardware, software, and infrastructures that makeup ICT.

"...to access, manage, integrate, evaluate and create information..."

Technology is used for an ever-increasing range of purposes to accomplish many different kinds of tasks. This phrase is meant to reflect that range as well as to define five critical components of ICT literacy. The five components represent a continuum of skills and knowledge and are presented in a sequence suggesting increasing cognitive complexity. After discussions regarding the kinds of tasks represented by each component, the panel agreed to the following definitions:

- Access—knowing about and knowing how to collect and/or retrieve information.
- Manage—applying an existing organizational or classification scheme.
- Integrate—interpreting and representing information. It involves summarizing, comparing and contrasting.
- Evaluate—making judgments about the quality, relevance, usefulness, or efficiency of information.
- Create—generating information by adapting, applying, designing, inventing, or authoring information.

"...in order to function in a knowledge society."

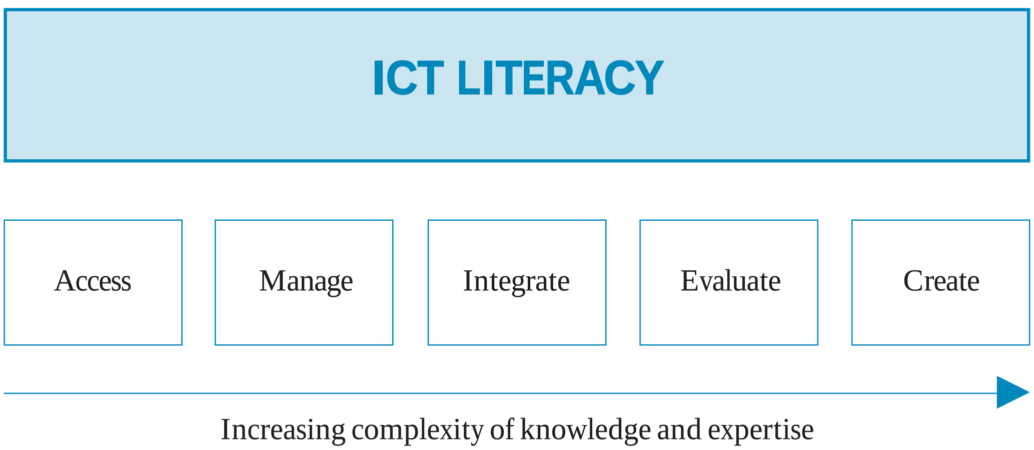
This phrase reflects the range of contexts in which individuals will be able to apply their ICT literacy—from defined ones such as graduating from school or functioning on a job to those which are less defined and less concrete but which can extend and enrich one's personal life. The phrase "in order to function" is meant to acknowledge the fact that ICT literacy will provide individuals with a means of contributing to and benefiting from economically developed or developing societies. We believe that ICT literacy skills are becoming increasingly important not only for nations to maintain or improve their standard of living but for the well being of individuals as well. The phrase "in a knowledge society" refers to the changing nature of cultures in the 21st century—an age in which ideas and information are increasingly the drivers of progress. The expanding roles of technology and access to information on a global scale have the potential to change, and hopefully improve, the way we live, learn and work.

4. Organizing the domain

Once we had defined what was meant by ICT literacy and laid out the assumptions underlying that definition, the next step was to develop an organizing framework for ICT literacy. This is an important step because the way in which the domain is organized affects test design and the kinds of tasks that will be developed to provide evidence about the status of ICT literacy in a population of interest. The panel's task was to define the critical organizing categories for the domain of ICT literacy and how they were related.

In our definition of ICT literacy, we identified five components we view as essential for functioning in a knowledge society: accessing, managing, integrating, evaluating and creating information in a technology context. These components, represented in Figure 1, formed the initial organizational scheme for the domain of tasks that make up ICT literacy.

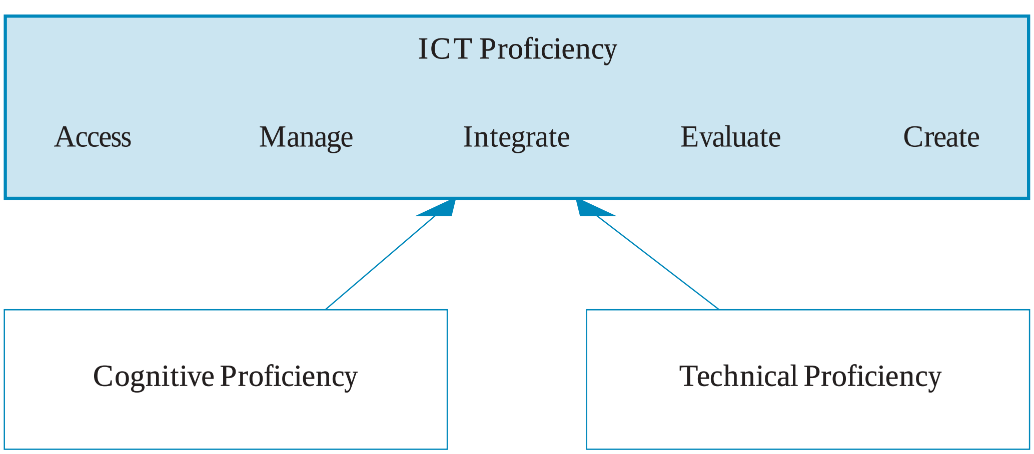
Figure 1



Upon further consideration, however, we chose to expand this unidimensional model to more fully represent the complexity of ICT literacy. This organizational scheme, shown below in Figure 2, illustrates the foundational set of skills and knowledge that underlie ICT literacy: cognitive and technical proficiency.

Figure 2

ICT LITERACY



The three proficiencies are defined as follows.

- Cognitive Proficiency—the desired *foundational skills* of everyday life at school, at home, and at work. Literacy, numeracy, problem solving, and spatial/visual literacy demonstrate these proficiencies.
- Technical Proficiency—the basic components of digital literacy. It includes a foundational knowledge of hardware, software applications, networks, and elements of digital technology.
- ICT Proficiency—the integration and application of cognitive and technical skills. ICT proficiencies are seen as enablers; that is, they allow individuals to maximize the capabilities of technology. At the highest level, ICT proficiencies result in innovation, individual transformation, and societal change.

As conceived in this framework, ICT literacy includes both cognitive and technical proficiency. For example, in order to successfully perform an ICT task such as searching the Internet to find and compare treatment options for a medical condition, an individual must apply reading and problem solving skills (cognitive) and be able to access information on the Internet using a search engine (technical). While cognitive and technical proficiencies are both necessary components of ICT literacy, each is a distinct domain. Cognitive and technical proficiency each represent independent domains in which the associated knowledge and skills interact to influence ICT literacy. An example is shown below in Figure 3.

Figure 3

	Low technical proficiency	High technical proficiency
High Cognitive Proficiency	A	B
Low Cognitive Proficiency	C	D

One would expect that individuals with low cognitive proficiency but high technical proficiency (cell 'D') would be able to perform particular technical tasks in which they had been trained. However, they would probably not possess the kind of generalizable skills or knowledge that could help them work with new applications or perform novel tasks and they would most likely not be able to acquire such skills independently. People with high cognitive proficiency but low technical proficiency (cell 'A') would require technical training (and possibly motivation or practice time) in order to develop ICT proficiency but would be expected to do so and once engaged with ICT would be able to acquire new skills and knowledge quickly and independently.

The representation of ICT literacy shown in Figure 2 provides an organizational scheme for both developing large-scale assessments or individual tests and evaluating existing measures. The framework leaves open the possibility that different constituencies could develop different assessments or individual tests for different purposes by focusing on various parts of the framework itself and by defining the kinds of evidence that might be associated with each. For the purposes of discussion, we present three types of assessments that might be developed using the ICT literacy framework: holistic, component and diagnostic. Each is described in turn below. Additional detail about possible assessment tasks can be found in Appendices A, B and C. These sample tasks are also available in a more interactive form at www.ets.org/research/ictliteracy/index.html.

A holistic assessment would be of most interest to constituencies who wished to focus on how well a test taker completed a given set of tasks rather than on the component skills that make up those tasks. An example would be a task that required test takers to create a flyer for a neighborhood clean-up day. Specific task requirements as well as information such as when and where the event was to be held would be provided and test takers would be scored on how well they completed the final product. (A more detailed illustration of this task can be found in Appendix A.) Such an assessment would allow one to rank order groups or individuals and to make general statements about their ICT knowledge and skills.

Alternatively, one might choose to build an assessment that independently measured the knowledge and skills associated with each of the five components of ICT Proficiency (access, manage, evaluate, integrate and create). This component assessment would result in a measure that could provide general information about the kinds of generative ICT tasks a population, or an individual, could perform. An example of one such task and its associated proficiencies is presented below. (For a more detailed illustration of this task, see Appendix B.)

Scenario: Employees in your company have been asked to provide information about the technology training courses they have taken during the past year. They have sent e-mail messages to their supervisors and each supervisor has forwarded the information to the director of human resources. You've been asked to organize the information, evaluate the extent to which company-based courses are being utilized, and make a recommendation to the human resources department about which courses should be continued next year.

Access	Select and open appropriate e-mails from inbox list.
Manage	Identify and organzie the relevant information in each e-mails.
Integrate	Summarize the interest in the courses provided by the company.
Evaluate	Decide which courses should be continued next year, based on last year's attendance.
Create	Write up your recommendation in the form of an e-mails to the vice president of human resources.

But if an individual or a group of individuals performed poorly on this measure, one would be hard pressed to understand or explain why. Were there underlying reading or language problems? Did test takers have sufficient technical knowledge to complete the tasks presented in the ICT measures? To understand what role these other domains contributed one would have to include cognitive and technical tasks in the assessment or test. Alternatively, one might want to focus on particular ICT proficiencies (for example, how well a person can access and manage information) and their underlying cognitive and technical components. This would involve creating tasks that measured these types of skills and knowledge across the three proficiency domains. These measures would provide evidence separating literacy and technology proficiencies from ICT proficiency. Such information would be useful for constituencies such as adult basic education centers interested in diagnosing and remediating problems students are having accessing information on the Internet. A series of tasks that might be appropriate in this context are presented below (and in more detail in Appendix C).

Scenario: Following a stroke, your mother has been diagnosed with an atrial septal defect, or a hole in one section of her heart. While not an emergency, her doctor has recommended open-heart surgery to repair the hole and reduce the risk of additional strokes. You would like to find several reliable sources on the Web that recommend treatment options for this condition.

Access	Using a search engine, locate sites that have articles about holes in the heart, or atrial septal defects.

Students having trouble with this basic ICT task could be presented with related cognitive and technical tasks to help diagnose what was causing their difficulty. For example, students might be presented with multiple-choice questions asking them to select the best word or phrase to use when searching for some specified information. Included among the choices might be terms that are overly general or specific. Students having difficulty with this type of task might need practice in defining categories and efficient search strategies. In addition, very basic computer tasks, such as opening a search engine, clicking on sites, and navigating back to the search engine from those sites, might uncover technical skills requiring review or training.

Currently, there are various measures of literacy, numeracy and problem solving being used in large-scale assessments of school age and adult populations. There is also a measure of technical knowledge and under-standing that is being used with school-aged populations. These are traditional paper and pencil measures. No attempt has been made, however, to build computer-based tasks to measure the integration of these cognitive and technical domains or to separate out the role each plays in the development of these more generative ICT proficiencies. The panel believes that the measurement of ICT literacy using paper and pencil will limit the ability to assess the full domain of knowledge and skills. Valuable information will be lost if assessment tasks are not embedded in real-world settings incorporating technology. For example, the measurement of an individual's ability to search for and access information would be hindered if the measurement did not provide an opportunity to log onto the Internet or a similar type of environment.

5. Next steps

As the panel began its deliberations about ICT literacy and how should it be defined and operationalized, we soon recognized that many of our discussions focused around the issue of the digital divide. This divide is commonly defined in terms of connectivity and the inequalities of access that exist both within and across countries. The more important issue the panel recognized was that the true potential of ICT—that is, the ability to transform individuals and societies—came not just from being wired together but also from having the knowledge and skills to use technology and to understand the roles it can play in our lives. As the president of Morris Brown College recently stated, “Merely having access to a box—an information box—does not necessarily mean that you have improved, or that you're more literate, or that you're better able to solve problems in the community” (Young, 2001).

This perspective led the panel to determine what they saw as the important issue facing us as society continues to invest in technologies and as technology continues to alter the way we work and live our lives. Then we wanted to use this storyline as a lead in to the definition of ICT literacy and how it should be operationalized into a framework. This report has taken the initial steps in building a framework by providing a consensus definition of ICT literacy and a model that can be used to further operationalize this construct for a variety of purposes.

The next steps will involve defining the kinds of evidence that should be gathered with respect to each level of the model—ICT, cognitive and technical proficiencies—and the kinds of activities that would elicit that evidence. This evidence and related activities will vary depending on the purpose of the planned assessment or test.

The framework begun with this paper, along with a prototype of online tasks, will allow ETS to discuss the potential for large-scale assessments or individualized tests with potential sponsors. The major stakeholders who will be interested in this framework and its resulting assessments are international and diverse, and therefore create a unique challenge as well as opportunity. They include government policy makers, corporate leaders, industry associations, unions, workforce groups, educators (K-12, higher education, national educational associations, researchers), consumer and public interest groups, and relevant international associations. The buy-in, cooperation, and support of these groups will be essential in the achievement of global ICT literacy.

ICT has become a permanent part of everyday life. It fundamentally changes how we live, learn, and work. Because ICT is considered an emerging and fundamental literacy, significant attention must be paid to insuring that all citizens have access and opportunity to gain the needed skills to function effectively in a knowledge society.

Appendix A

Sample assessment task—ICT proficiency

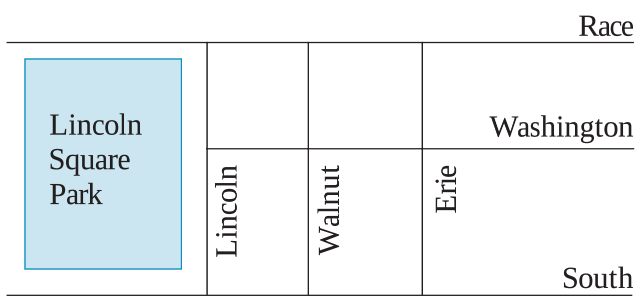
- Holistic assessment of ICT skills and knowledge
- Scenario presented along with a variety of tools (spreadsheet, word processor, etc.)

In this type of assessment, test takers would be evaluated solely on the end product they created (for example, a database, presentation, or document). Component skills would not be isolated and individually assessed. Instead, a scoring scheme would be developed which defined levels of performance and the criteria for reaching each level. This scheme would represent the collective judgments of experts in the field about what adults should know and be able to do in the ICT domain. Below is an example of what one task in a holistic assessment might look like. A complete assessment would include a number of different tasks that vary in difficulty and require a range of ICT knowledge and skills.

Opening scenario (Community context)

You've volunteered to create a flyer for a community clean-up day to be held in your neighborhood. Include the map below along with the following information and create an attractive one-page flyer for the event.

The event will take place on Saturday, May 6th from 1:00 until 4:00. Volunteers are being asked to meet at Lincoln Square Park. Event organizers would like a tear-off registration slip to be included on the flyer where volunteers



can print their name, address and phone number. The registration forms should be dropped off at the community center on Race Street by May 1st.

To complete this task, test takers would need to use a word processing program to create a flyer. The final product would be scored on the accuracy and completeness of the information it contained (e.g., did the flyer include all the relevant information about dates and times, the map and the tear-off registration form?). Additional scoring points might include evaluating the layout and inclusion of graphic elements (borders, lines, etc.).

Appendix B

Sample assessment task—ICT components

- Focus on the components of ICT proficiency: access, manage, integrate, evaluate, and create
- Present a scenario followed by tasks addressing each of the five components

Below is an example of what one task in this type of assessment might look like. A complete assessment would include a number of different tasks that vary in difficulty and require a range of knowledge and skills in technical, cognitive, and problem-solving domains.

Opening scenario (Workplace context)

Employees in your company have been asked to provide information about the technology training courses that they have taken during the past year. They have sent e-mail messages to their supervisors and each supervisor has forwarded the information to the director of human resources. You've been asked to organize the information, evaluate the extent to which company-based courses are being utilized and make a recommendation to the human resources department about which courses should be continued next year.

The five components

Based on this scenario, test takers would be presented with a series of tasks. Each task would be designed to measure one of the five components, as summarized in the chart below. While the sequence in which individual test takers undertake these tasks might vary, each component could be scored discretely in order to better understand its relative contribution to an individual's overall ICT proficiency.

Access	Select and open appropriate e-mails from inbox list.
Manage	Identify and organzie the relevant information in each e-mails.
Integrate	Summarize the interest in the courses provided by the company.
Evaluate	Decide which courses should be continued next year, based on last year's attendance.
Create	Write up your recommendation in the form of an e-mails to the vice president of human resources.

Test takers might work from a screen that presents all of the task components and allows them to select the order in which they complete those tasks. An alternate approach would be have test takers work through a structured series of tasks with the first component presented, followed by the second component and so on.

A more detailed description of the component tasks is presented below.

Access and manage task

Task Description: *Seven supervisors have sent information about training courses to Ann Simpson, Director of Human Resources, and she has forwarded them to you. Find and open each of those e-mails in your inbox. Select the text from each e-mail that provides information about training course attendance and copy it all into a single file.*

Test takers would be presented with a simulated inbox, similar to the sample shown below. Some might chose to open all the e-mails and then select the relevant information. Others might open one e-mail, select the critical information and then move on to the next. Whatever the sequence, to complete the task correctly test takers would be expected to open each of the correct e-mail messages and paste all the relevant information into a file.

Sample In-Box:

From	Subject	Received	Size
Simpson, Ann	FW: Training	12/17/01 10:32 AM	3 KB
Simpson, Ann	FW: Course Information	12/17/01 10:44 AM	2 KB
Davidson, Denise	RE: Lunch	12/17/01 10:57 AM	7 KB
Simpson, Ann	Work Objectives	12/17/01 11:11 AM	5 KB
Simpson, Ann	FW: Classes Taken	12/17/01 11:27 AM	3 KB
Corporate Communique	Virus Alert	12/17/01 12:01 PM	4 KB
Simpson, Ann	FW: Courses This Year	12/17/01 12:15 PM	4 KB
Simpson, Ann	FW: Training Classes	12/17/01 12:49 PM	2 KB
Gonzalez, Frank	Team meeting	12/17/01 1:08 PM	8 KB
Simpson, Ann	FW: Thursday Staff Meeting	12/17/01 1:11 PM	3 KB
Simpson, Ann	FW: Training Courses	12/17/01 1:59 PM	2 KB
Salverston, Amy	RE: Phone Billing	12/17/01 2:14 PM	6 KB
Mirano, Leslie	Training Class Question	12/17/01 2:48 PM	5 KB
Jenkins, Ralph	Update	12/17/01 3:19 PM	3 KB
Simpson, Ann	Memo for Davidson	12/17/01 3:21 PM	4 KB
Ellis, Edward	Re: Phone Conference	12/17/01 3:56 PM	2 KB
Simpson, Ann	FW: Staff Training Courses	12/17/01 4:17 PM	2 KB
Rogers, Charlie	FW: Memo Format	12/17/01 4:45 PM	3 KB

Ann – Jason and I met yesterday and have a schedule for the next team meetings. We will send that information out to everyone later today. Here is the information you requested about training courses. In my area, 25 people took one or more training classes this year. 15 people took Learning Excel, Level 1 (March 27 and 28), 20 took Introduction to Outlook (June 3 and 4) and 5 took Flash, Level 2 (October 19 and 20). The first two courses were given on site and the last was at the community college. We have gotten particularly positive feedback about the Outlook course. Let me know if you need any additional information.

- E. O'Brien

Integrate

Task Description: *You want to look at all the information the supervisors have provided so that you can see which of the courses taught at the company were most popular. Represent that information in a way that will help you make the recommendation about which courses to continue next year.*

Test takers would need to decide the best way to integrate and compare the information they have selected in the previous task. They might present the information in a list or series of lists, in a table, etc. In the sample response shown below, the information from the seven e-mail messages has been used to create a table that allows one to quickly compare course location and attendance across courses.

Sample response:

Location	Name of course	Number of employees
On site	Learning Excel	31
On site	Introduction to Outlook	50
On site	Visual Basic	5
On site	HTML	25
On site	Networking Essentials	2
Advantex Computer Training	C++	5
Community college	Flash, Level 2	5
Community college	Windows NT	17

Evaluate and create

Task Description: *Using last year's attendance figures for courses offered by the company, decide which courses should be offered next year. Write an e-mail to Ann Simpson with your recommendation, including as attachments any tables or charts that support your position.*

Test takers would need to identify the on-site courses with the best attendance based on the supervisor's reports. They would then to write up their recommendation and attach supporting documentation. Scoring models would be created to focus on the skills and knowledge deemed most relevant to assess for a particular population. For example, one might be interested in the extent to which test takers were able to support their recommendation with evidence from the original supervisor's e-mails, the sophistication of supporting documentation, or the test taker's ability to use software to create tables or graphs.

Appendix C

Sample assessment task—Diagnostic assessment

This type of assessment would allow one to investigate the cognitive and technical proficiencies underlying particular ICT components. On the surface, the Diagnostic Assessment would look exactly like the assessment of ICT Components. Only if and when test takers had difficulty with a component task would they see new types of tasks designed to assess underlying cognitive and technical skills.

The results of this kind of assessment could be used in a variety of ways:

- The assessment could provide an overall score of a person's ICT, cognitive and technical proficiency.
- A more detailed score reporting system might be developed that profiled specific strengths and weaknesses that an individual demonstrated.
- Links to existing or specially developed instructional materials could be provided to help teachers in education or training settings.
- Based on a person's performance, targeted instructional goals and suggestions on how best to reach those goals could be made available.

Just like in the ICT Component assessment, a number of scenarios in different contexts would be presented. One sample scenario, developed in a health context, is shown below.

Opening scenario (Health context)

Following a stroke, your mother has been diagnosed with an atrial septal defect, or a hole in one section of her heart. While not an emergency, her doctor has recommended open-heart surgery to repair the hole and reduce the risk of additional strokes. You would like to find several reliable sources on the Web that recommend treatment options for this condition.

The five components

Based on this scenario, test takers would be presented with a series of tasks organized around the five components, as summarized in the chart on the following page.

Access	Using a search engine, locate sites that have articles about holes in the heart, or atrial septal defects.
Manage	Evaluate the sites and identify three that you would expect to provide reliable medical information.
Integrate	Identify the treatment information in each article.
Evaluate	Compare and contrast the treatment options suggested in the articles.
Create	Devlop a Word document with treatments listed (citing sources) to share with physician.

As each task was completed it would be automatically scored. If a test taker did not complete a task correctly, related cognitive and technical tasks would be presented to try and determine if one or both of those areas were contributing to the individual's difficulty.

An example of how the Access task might be broken down follows.

Access task

Task Description: *Use the search engine provided to find three sites with information about your mother's medical condition as described in the opening scenario.*

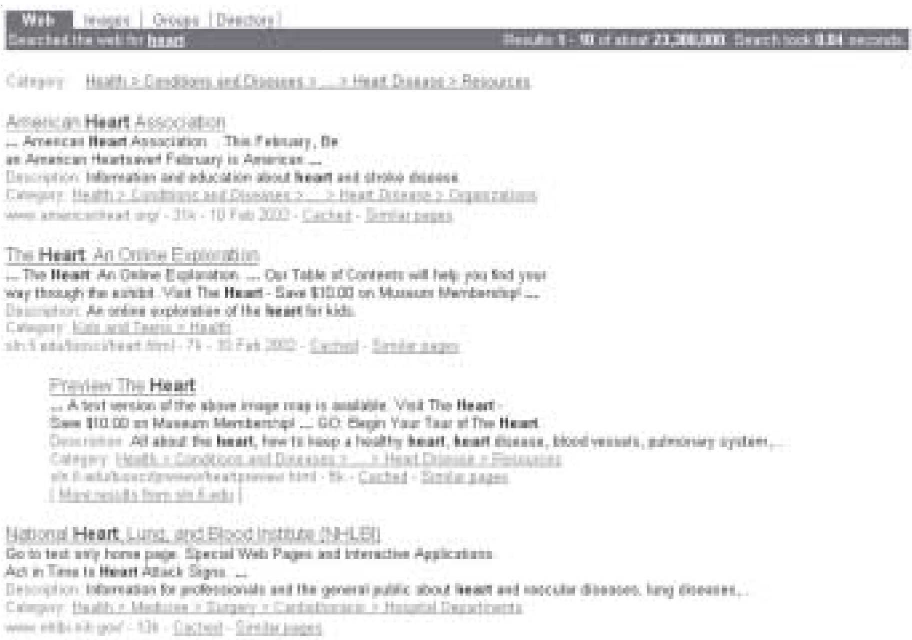
Assessing underlying technical skills

If a test taker did not complete the task correctly, one question would be whether he or she had the requisite technical skills. Technically, this access task requires test takers to open up a browser, type a word or phrase into the text entry box, and click on the Search button. They might additionally need to open a site and then navigate back to search engine. As an individual test taker completed this task, the computer would record clicks, typing and other actions. Based on the test taker's responses, additional discrete computer-based tasks might be presented (e.g., "Type the phrase 'Movie Listings' into the search box" or "Click on the button that will take you back to the search page") to assess the technical skills underlying this basic Access task.

Assessing underlying cognitive skills

Cognitively, this access task requires a test taker to select or invent a search term that would yield the requested information. Some of the words or phrases in the task description and scenario would be more likely than others to provide the information needed. For example, typing in the phrase "hole in the heart" in one browser would yield the results shown below, none of which would be likely to include the information needed.

Typing in the more general term, "heart," would result in the following types of sites.



The word or phrase test takers used for their query would provide the basis for scoring this task (with a more precise phrase such as “atrial septal defect treatment” resulting in a higher score than “treating heart”). The program might also track if test takers refined their search based on the results each search yielded. If a test taker did not perform well on this task, other less open-ended tasks might be presented. These might include multiple-choice questions that asked test takers to select from the choices provided the best term to search for specified information. Questions that focused on general versus specific categories might provide additional diagnostic information. Another possibility would be to present a similar search task that was not computer based, such as locating specified information in the Yellow Pages and seeing if the test taker could successfully complete that task. The goal of any of these or additional follow-up tasks would be to try and identify underlying areas of difficulty that might be contributing to poor performance on the computer search task.

References

American Society of Training and Development (ASTD) *State of the industry report.* (2000). Alexandria, Virginia. Author.

Benton Foundation. (2000). *The e-rate in America: A tale of four cities*. Washington, DC: Author.

Blau, Andrew. (2001). *More than bit players: How information technology will change the way nonprofits and foundations work and thrive in the information age*. New York: Surdna Foundation.

Bollier, D. (2000). *Ecologies of innovation: The role of information and communications technologies: A report of the eighth annual Aspen Institute Roundtable on Information Technology*. Washington, DC: The Aspen Institute.

Cline, W. (1997). *Trade and income distribution*. Washington, DC: Institute for International Economics.

Commission on Technology and Adult Learning. (2001). *A vision of e-learning for America's workforce*. Alexandria, VA, and Washington, DC: American Society for Training and Development and the National Governors Association.

Committee on Information Technology Literacy. (1999). *Being fluent with information technology*. Washington, DC: National Academy Press.

de los Santos, G.E., de los Santos Jr., A.G., and Milliron, M.D. (Eds.). (2001). *Access in the information age: Community college bridging the digital divide*. Mission Viejo, CA: League for Innovation in the Community College.

Digital Opportunity Initiative. (2001). *Creating a development dynamic*. New York, NY: Accenture, Markle Foundation, United Nations Development Programme. http://www.opt-init.org/framework.html.

Educational Development Center. (2000). *IT pathway pipeline model: Rethinking information technology learning in schools*. Newton, MA: Author.

Ellis, Carlene. (2001). *Innovation in education: The increasing digital world—issues of today and tomorrow*. Presented at the National IT Workforce Convocation of the Information Technology Association of America, San Diego, CA. http://www.intel.com/education/community/contribution.

Goslee, S. (1998). *Losing ground bit by bit: Low-income communities in the information age*. Washington, DC: Benton Foundation.

Hart, D. (2001). Bridging the digital divide, *Hemispheres*.

Information Technology Association of America. (2001). *When can you start? Building better information technology skills and careers.* Arlington, VA: Author.

Information Technology Association of America. (2000). *Bridging the gap: Information technology skills for a new millennium*. Arlington, VA: Author.

Katz, S.N. (2001, June 15). In information technology, don't mistake a tool for a goal. *The Chronicle of Higher Education, The Chronicle Review, 47*(40), B7-B9. http://www.chronicle.com/ weekly/v47/i40/ 40b00701.htm.

Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured* (ETS RR-01-25). Princeton, NJ: Educational Testing Service.

Morino Institute. (2001). *From access to outcomes: Raising the aspirations for technology initiatives in low-income communities*. Reston, Virginia: Author.

National Association of State Boards of Education. (2001). *Any time, any place, any path, any pace: Taking the lead on e-learning policy*. Alexandria, VA: Author.

National Telecommunications and Information Administration. (2000). *Falling through the net: Toward digital inclusion*. Washington, DC: The US Department of Commerce. Author.

National Telecommunications and Information Administration. (1999). *Falling through the net: Defining the digital divide*. Washington, DC: The US Department of Commerce. Author.

National Telecommunications and Information Administration. (1998). *Falling through the net II: New data on the digital divide*. Washington, DC: The US Department of Commerce. Author.

National Telecommunications and Information Administration. (1995). *Falling through the net: A survey of the "have nots" in rural and urban America*. Washington, DC: The US Department of Commerce. Author.

Organisation for Economic Co-operation and Development. (2001a). *Understanding the digital divide*. Paris: Author.

Organisation for Economic Co-operation and Development. (2001b). *The well-being of nations*. Paris: Author.

Organisation for Economic Co-operation and Development. (2000b). *Schooling for tomorrow: Learning to bridge the digital divide*. Paris: Author.

Organisation for Economic Co-operation and Development. (1999). *The economic and social impact of electronic commerce: Preliminary findings and research agenda*. Paris: Author.

Organisation for Economic Co-operation and Development and Human Resources Development Canada (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey.* Paris: Author.

Organisation for Economic Co-operation and Development and Statistics Canada (1995). *Literacy, economy and society: Results of the first International Adult Literacy Survey.* Paris and Ottawa: Author.

Organisation for Economic Co-operation and Development and Statistics Canada (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey.* Paris and Ottawa: Author.

Reinhardt, A. (March, 1995). New ways to learn. *Byte Magazine*. http://www.byte.com/art/9503/sec7/ art1.htm.

Venezky, R.L. (2001). *Assessment of ICT concepts and skills: Summary and recommendations for PISA*. Unpublished manuscript, University of Delaware.

Young, J. (2001, November 9). Does the digital divide rhetoric do more harm than good? *The Chronicle of Higher Education, Information Technology.* http://chronicle.com/infotech.