

**THE EVALUATION OF STUDENTS
IN THE CLASSROOM: A HANDBOOK AND
POLICY GUIDE**

Second Edition

**Department of Education
Division of Evaluation, Testing & Certification
St. John's, Newfoundland**

September 2000

Table of Contents

Table of Contents	i
Acknowledgments	iii
Preface	1
Understanding Evaluation	2
Meaning of Evaluation	3
Philosophical Bases of Evaluation	4
Summary	5
The Evaluation Process	6
Types of Evaluation	7
Outcomes	10
Process and Product in Instruction and Evaluation	14
Differentiated Evaluation	17
Evaluation Issues in Direct Instruction	20
Summary	21
Sources of Evaluation Data	22
Learning Journals	24
Observation	25
Performance-Based Assessment	27
Portfolios	28
Projects	30
Self-Evaluation Reports	33
Teacher-Student Conferences	34
Tests	35
Work Sample Files	42
Summary	42
Interpreting and Using Evaluation Data	43
Grading	44
Reporting	47
Promotion	51
Sample Policy	52
Appendix A	66
Appendix B	70
Appendix C	71
Appendix D	73
Appendix E	74
Appendix F	75
Appendix G	76
Appendix H	77
Appendix I	78
Appendix J	79
Appendix K	80
Appendix L	81

Appendix M	83
References Cited	87

ACKNOWLEDGMENTS

The Department of Education wishes to acknowledge the efforts of all those individuals who provided input and feedback during the development of this document. In particular, the contribution of the following people, who served as members of the committee that oversaw the project, is gratefully acknowledged:

Russ Blagdon (Chair)
Frank Crews
Joyce Fewer
Edith Furey
Ivan Hibbs
Susan Oke-McCarthy
Diane Vey-Morawski
Pamela Williams

PREFACE

Student achievement is profoundly influenced by the evaluation practices used by teachers in the classroom. Research in education has consistently shown achievement monitoring and feedback to be extremely important factors in the instructional process which have a significant impact on student success. Instruction can seldom be effective without a comprehensive evaluation plan that is carefully and systematically implemented in the classroom.

This handbook has been written to help provide classroom teachers, and others who provide direct support to classroom instruction, with information on good classroom evaluation practices. While it has a theoretical basis supported by references to evaluation literature, it is intended for practical use in policy development and in classroom application.

Evaluation is much more than testing. It is a continuous and comprehensive process, rather than a series of sporadic and independent events. Evaluation should guide student learning daily rather than simply provide information for making decisions on promotion at the end of the year. In this document, outcomes are highlighted, since they represent the beginning and the end of the instructional process. It is recognized that the process students go through in coming to know is just as important as what they know. It is noted that comprehensive evaluation will rarely occur when only one source is used to gather evaluation data, and if certain students are not to be at a disadvantage, special arrangements should be made in order to evaluate their achievement. It is pointed out that students should be graded in relation to the outcomes that have been set for them, not merely in relation to other students. Communication is emphasized since information must flow continuously between the home and the school. Decisions on promotion or grade retention must always be based on a consideration of what is best for students, not only for the next year, but for the rest of their lives.

The policy statements and guidelines which have been extracted from discussions on student evaluation could serve as a guide for any district wishing to modify, develop or implement a policy on student evaluation.

UNDERSTANDING EVALUATION

Evaluation of student achievement is a crucial and pervasive issue of concern in every classroom. Traditionally, evaluation activities were used to certify or promote students. Currently, evaluation is recognized as a comprehensive process, not just something that occurs only at the end of some learning episode. It is an integral part of the teaching-learning process. Evaluation activities provide information for the identification of students' strengths and needs, a basis for the assessment of teaching strategies and data for important decisions regarding corrective or enrichment activities.

If a classroom is to function effectively, a well-planned system of evaluation is essential. Without such a plan, it is difficult to determine the abilities of students and to decide which methods of instruction are most appropriate for a particular class or group of students. It can have serious consequences for the teacher, even though the effects are not personal and only temporary. For the student however, the effects of evaluation activities can be more serious and long-lasting.

In this chapter, the differences between evaluation and measurement are noted and the philosophical foundation for a comprehensive and effective evaluation policy is provided. While many principles and beliefs are implicit in the philosophy, more precise and explicit statements that should guide the evaluation process are provided.

Meaning of Evaluation

The term *evaluation* is frequently interchanged with the terms *assessment*, *measurement* and *testing*. These terms have distinct meanings. In the Standards for Teacher Competence in Educational Assessment of Students developed by the American Federation of Teachers, the National Council on Measurement in Education and the National Education Association in 1990, assessment is defined as the process of collecting information to be used in making educational decisions about students, to give feedback to students, to make judgements regarding curriculum and instruction and to inform policy. Nitko (1996) defined measurement as a procedure for assigning numbers to a specified attribute in such a way that the numbers indicate the degree to which a person possesses the attribute. He also defined a test as an instrument or systematic procedure for gathering measurements. Gronlund (1985) stated that evaluation puts a value upon assessment results concerning their desirability. It seeks to answer the question "of what value?" Evaluation is a more comprehensive and inclusive term than assessment, measurement, or testing. He also pointed out that evaluation may involve quantitative descriptions (measurement) and qualitative descriptions (non measurement) of pupils, where the main emphasis is "on the extent to which learning outcomes are achieved"(p.6). Both types of evaluations and descriptions are necessary for comprehensive student evaluation.

Evaluation is, essentially, a process involving a value judgement regarding the achievement of stated outcomes. It is a comprehensive process encompassing the entire range of students' activities and experiences. It is continuous in that there is no starting, no finishing point.

Philosophical Bases of Evaluation

Education is concerned with the integral formation of students. Hence, it should be directed toward their physical, emotional, aesthetic, social, moral, intellectual and spiritual development.

Outcomes designed to guide evaluation should take direction from this multidimensional development. Any concept of human knowing takes in the whole of human life: the external senses, the imagination, the feelings and emotions, the ability to think and reason and to make responsible decisions. This has implications for teaching and learning and for evaluation. Outcomes should be grounded in a theory of knowing which recognizes the multidimensional nature of human beings. There are outcomes which deal with facts, critical thinking, values, attitudes, appreciations, and beliefs. There are outcomes concerned with study skills, group skills, and the many manipulative and manual skills that form part of the school curriculum. Any system of evaluation has to be based on who students are and how they know and act as learners.

The following principles of evaluation should be evident from the philosophy. They are more fully developed in various sections of this document.

- Evaluation takes its direction from a definition of education which, stated in its broadest sense, is to enable students to realize their potential as human beings.
- Evaluation is a means to an end and never an end in itself.
- Evaluation, by definition, connotes value.
- Evaluation involves teacher judgement.
- Measurement implies the use of some tool or instrument such as a test or scale and provides a quantitative description of observed phenomena.
- Validity and reliability are of paramount concern in any evaluation activity.
- Evaluation is continuous in that it is taking place all the time and comprehensive in that it is concerned with the cognitive, psychomotor and affective domains of human development.
- Evaluation is an integral part of the teaching-learning process.
- The validity and reliability of evaluation improve when multiple sources of evaluation data are used.
- Every evaluation should be made with reference to specified outcomes.
- Students and parents should be aware of the outcomes of educational programs.
- Students should play an active role in their evaluation.
- Evaluation procedures should take into consideration individual differences among students.
- Evaluation procedures are important for the guidance of individual students.

- Evaluation of students involves more than a single appraisal at any one time.
- Self evaluation is very beneficial to students.
- The process of evaluation begins with the outcomes of the educational program.
- The primary purpose of evaluation is to determine the extent to which students have achieved the intended outcomes of the educational program.
- Evaluation gives a picture of where students are, how they are progressing, and what their needs are.
- Evaluation is concerned with some attributes that are quantitative and some that are qualitative.
- Evaluation is a more inclusive concept than assessment - assessment involves the collection of data needed for evaluation, but it is not evaluation per se.

Summary

Student evaluation is a process basic to teaching. Evaluation is not an add-on feature of instruction but an integral part of it, since the information it provides allows teachers to make adjustments to outcomes and teaching methodologies.

Evaluation must be student-centered where the process and subsequent decisions reflect a genuine concern for the interest of each student. This can best be achieved by following the basic philosophy and principles outlined, and by recognizing that evaluation activities are designed to help a student.

Teachers have a responsibility to provide accurate, reliable and justifiable evaluations which do not undermine the integrity, self-confidence, or enthusiasm of students. Comprehensive evaluation involves the use of a variety of sources and is referenced to specific learning outcomes. Its implementation ensures that the total growth of every student is recognized.

THE EVALUATION PROCESS

The evaluation of student learning does not take place in a vacuum. It takes place in the context of teaching and learning, and many factors in this context have implications for student evaluation. Figure 1 shows the dynamic relationship among outcomes, instruction and evaluation. Teaching and learning begin with intended outcomes. Teachers determine the strategies to be used to achieve these outcomes and use evaluation to determine the extent to which the outcomes have been achieved. The arrows in the diagram indicate that experience in either of the three spheres of activity can result in adjustments being made in either of the other two spheres. In Newfoundland and Labrador, however, a centralized curriculum exists, and changes to outcomes can be made only where an Individual Support Services Plan (ISSP) has been developed.

In this chapter, the different types of evaluation and outcomes are considered as well as process and product approaches to education, differentiated evaluation and the effective teaching research.

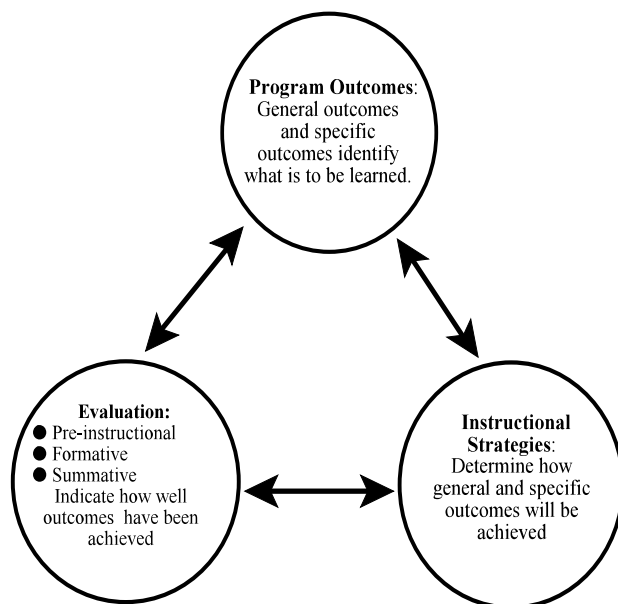


Figure 1. The Teaching-Evaluation Process

Types of Evaluation

For the purpose of this discussion, three types of evaluation are defined: pre-instructional, formative, and summative. Each type has a different purpose. To any observer, one type of evaluation looks much like another because the type of evaluation is determined by how the results are used. If results are used to determine appropriate starting points for instruction, evaluation may be regarded as pre-instructional. When data are used to determine how well instruction is going, evaluation is formative. If the purpose is to determine the extent to which students have achieved the outcomes of the unit of study, evaluation is summative. However, where it is decided after a summative evaluation that students have not achieved outcomes to a satisfactory degree, and a decision is made to reteach, evaluation intended as summative becomes formative. Whether pre-instructional, formative or summative decisions are being made, they should be based on data from a variety of sources.

Pre-Instructional Evaluation

In the pre-instructional phase of evaluation, an attempt is made to determine what each student brings to the learning situation or where that student stands in relation to stated outcomes. In a province where there is a centralized curriculum, as is the case in Newfoundland and Labrador, the teacher can determine, in a general sense, where the student stands by examining the curriculum. Where a student has special abilities or special needs, this would have been noted by previous teachers, and appropriate communication would have been placed in the student's cumulative file.

While appropriate information is necessary to adapt instruction and curriculum properly, caution must be exercised to prevent the development of improper expectations for student achievement. Pre-instructional evaluation should be seen as an indication of what a student's strengths and needs are, not as a measure of what they are capable of doing. Where low expectations are a result of pre-instructional evaluation, a vicious circle can develop. As students fall further and further behind, teachers expect less and less, and as teachers expect less and less, students fall further and further behind.

It might seem the safest thing for a teacher to do is hold high expectations for all students. However, some students are capable of achieving at a higher level than others. It is unrealistic to assume everyone can achieve at the same level. An appropriate expectation for all students would be one that sees them working to their maximum and progressing at a steady rate.

Pre-instructional evaluation serves a number of purposes. It provides information on the level of readiness of each student in relation to outcomes. It may also indicate the learning styles of students as well as the need for particular adaptations to instruction.

There are a number of sources available for pre-instructional evaluation. These include cumulative records, standardized and teacher-made tests, meetings with parents and teachers, and teacher-student conferences. Teachers should use a variety of these sources to acquire needed information, continuously updating the information during the teaching-learning process.

Formative Evaluation

Formative evaluation is concerned with how things are going. It is evaluation which occurs during the teaching-learning process rather than before or after it. This is the most valuable type of evaluation, concerned with both the processes and products of learning. Its fundamental purpose is the identification of a student's strengths and needs with respect to specific outcomes so that necessary adaptations and modifications can be made as soon as possible. The diagnostic aspect is important in formative evaluation where emphasis is on providing the most suitable program of instruction or identifying and eliminating barriers to success. Early intervention, advanced placement, or remedial assistance should be provided for the learner when the need for additional support is indicated. Standardized assessment instruments such as tests are often an essential component of comprehensive assessment used in the identification of exceptionalities. A discussion of various standardized assessment tools is contained in Programming for Individual Needs: Assessment and Evaluation (Newfoundland Department of Education, 1998).

A systematic process for monitoring student progress is required in formative evaluation. Results of this type of evaluation should be provided immediately to the student. The benefit of continuous feedback to the learner is well documented in the research on effective teaching. Communication with the student is a key element of formative evaluation. It is also important for parents to be well informed about the progress of instruction. All parents want the best for their children, and they have a right to know how their children are doing as opposed to how they have done. When parents know how students are progressing, they can take an active role in helping create a good learning environment.

An underlying assumption of formative evaluation is that in learning, the responsibility for success should be shared. Principals, teachers, parents and students generally agree on the purposes of schooling, and all are dedicated to the proposition that children can and will learn. Parents should not assume their children's education is the complete responsibility of the teacher. Students themselves have to assume the major share of responsibility for their own learning. Parents have an obligation to encourage their children to do well and teachers should provide effective instruction, but in the final analysis, students are responsible for their own learning.

Although the most important source of formative evaluation data is observation, other sources can be used as well. Formative evaluation can be conducted by the teacher, or by students themselves as they evaluate their own or their classmates' work. It indicates the strengths and needs of the students, the program and the teaching methodology. As well, it provides work samples that are useful as part of the reporting process. When students are taught to evaluate themselves, it strengthens their sense of responsibility for their own learning. They also become aware of outcomes as they develop the capacity to evaluate their own work. As well, this type of evaluation can provide feedback to the teacher on student reactions to various aspects of the curriculum. Peer evaluation provides students with responses to their work other than the teacher's and it gives students an opportunity to compare their work with that of other students.

Summative Evaluation

Summative evaluation is used to grade, certify or select students. It is not intended to improve current instruction for the benefit of those being evaluated. It measures what has happened, not what is happening. But results can serve to indicate areas of strength and needs, and these results can be used to influence later instruction. In that sense, summative evaluation can have pre-instructional and formative dimensions.

Summative evaluation is the evaluation of student achievement of specific outcomes at the end of instruction. It is directed toward a general assessment of the degree to which outcomes are attained over the entire course or some substantial part of it. The essential characteristic is that a judgement is made about a student with regard to what learning has taken place or is supposed to have taken place. Summative evaluation is used to grade or certify students after instruction has taken place. Summative evaluation is used for diagnosis as well when deciding who should advance to the next level. It may be used to determine the present status of students, to determine eligibility for special services, to identify factors which are responsible for and influence growth and development, and to determine potential for future growth and development.

Data for use in summative evaluation can be obtained from a variety of sources. At the primary and elementary levels, teacher observation is a major source, while testing is used more frequently at the intermediate and senior high levels. In any event, the primary purpose is to gather sufficient data to ensure a valid judgment about a student's achievement of stated outcomes. Where outcomes cannot be appropriately evaluated using paper-and-pencil tests, other sources of evaluation data such as work samples, interviews, and observations can function as summative instruments if used separately for this purpose. Instruments used in summative evaluation should reflect the teacher's methods of instruction and the significance and sequence of the outcomes. The student's evaluation should be referenced to the outcomes of instruction rather than simply to relative rank among those evaluated. The instruments used should be flexible enough to permit a variety of student responses, and should provide scope for the demonstration of divergent thinking consistent with the outcomes of instruction. For evaluation having a summative component, students should be informed in advance of the event as to its timing, method and outcomes to be evaluated.

The tendency to average formative results according to a rigid, predetermined formula for summative purposes should be avoided. While the practice of 'counting' everything a student submits toward a course grade improves the sampling, it can also result in an unrealistically low appraisal of the student's knowledge in relation to the outcomes of instruction, resulting in a lessening of validity. The practice also penalizes the learner who achieves the outcomes at a later stage of the instructional program. It is important students have the freedom to err and profit from mistakes without fear of penalty or reprisal.

The best person to evaluate the progress of a student is the person responsible for instruction. It is the summative phase of evaluation which calls for the most careful exercise of the teacher's professional judgment.

Outcomes

Outcomes are goals defined in terms of human thoughts and performances. They describe what students are expected to demonstrate upon completion of an organized program of instruction. They spell out actions and values deemed to contribute to the individual and to society. Of paramount importance is respect for the worth and integrity of all students with all being given an equal opportunity to grow and develop to the fullest extent of their capabilities. A logical step in discussing evaluation is to examine the role of outcomes in the teaching-learning process.

Education involves learning as an active process through which learners continually construct or reconstruct meaning in the light of new experiences. Currently, education in Newfoundland and Labrador is guided by the Essential Graduation Learnings, which are broad, general statements describing student learning in terms of knowledge, skills and values. These learnings are cross-curricular and are supported by more specific outcomes in each area of study. The achievement of specific outcomes enables students to achieve the Essential Graduation Learnings.

Outcomes provide the basis for the development of a coherent curriculum. They convey instructional intent and the knowledge, skills and values students are expected to develop as a result of their learning experiences. That is not to say all actions in the classroom must be guided by outcomes. Eisner (1985) wrote about expressive outcomes. Such outcomes describe an educational encounter. Students work with a problem or engage in a task, but what is to be learned from the encounter does not always have to be specified. While outcomes provide guidance for learning, they should not stifle it. Students should always be free to explore or focus on issues of individual interest.

Curriculum Outcomes Structure

For the K-12 education system in Newfoundland and Labrador, outcomes are organized as Essential Graduation Learnings, General Curriculum Outcomes, Key Stage Curriculum Outcomes, and Specific Curriculum Outcomes.

Essential Graduation Learnings are general, cross-curricular descriptions of what a student's level of education should be by the end of Grade 12. General Curriculum Outcomes describe what students are expected to know, be able to do, and value upon completion of an area of study such as a particular discipline. Key Stage Curriculum Outcomes describe what students are expected to know, be able to do and value in an area of study at the key stages of the end of Grade 3, the end of Grade 6, the end of Grade 9, and the end of Grade 12. Specific Curriculum Outcomes describe what students are expected to know, be able to do, and value in an area of study at a particular grade level.

The day-to-day instructional program in an area of study is guided by the Specific Curriculum Outcomes for the year's work which, over successive years, lead to acquisition of key stage and general curriculum outcomes, and ultimately the essential graduation learnings.

Curriculum Implementation

The provincial curriculum is supported by authorized learning resources. While curriculum guides and learning resources usually suggest an organization for teaching and learning, teachers have the responsibility for organizing lessons on a short-, medium- and long-term basis. Lesson plans should be organized in consideration of the specific curriculum outcomes such that learning experiences are structured to achieve these outcomes. In effect, specific curriculum outcomes guide teaching, learning and the evaluation of student progress. Teachers should never lose sight of the fact that specific outcomes contribute in a very meaningful way to the achievement of broader outcomes.

As Gronlund (1995) noted, outcomes contribute to learning by providing direction in the selection of subject matter, teaching methods and materials, by guiding assessment, and by informing others about instructional intent. Outcomes are important to students. It is important for students to be informed of the expectations for their performance. Specific curriculum outcomes should be shared with students in advance of instruction. These outcomes should be written in student language.

Gronlund (1995) identifies a two-step procedure for clarifying outcomes. Once a curriculum outcome is stated in somewhat general terms, teachers should clarify the outcome by listing a sample of the types of performance to be accepted as evidence of the attainment of that outcome. This two-step procedure is illustrated as follows:

Step 1: State a general outcome such as *Students will demonstrate understanding of terms.*

Step 2: State specific outcomes such as:

Students will define the terms in their own words.

Students will identify the meaning of the terms used in different contexts.

Students will distinguish between and among terms which have similar meanings.

There is an advantage to stating outcomes in this way. While instruction is directed at achieving the specific outcomes, the teacher is very much concerned with the broad outcome. Specific outcomes do not become ends in themselves.

Educational outcomes have been described according to various classification systems or taxonomies. Bloom (1956) developed a taxonomy for the cognitive domain, Krathwohl, Bloom and Masia (1964) developed one for the affective domain and Simpson (1972) developed one for the psychomotor domain. Similar frameworks for classifying outcomes have been suggested by Gagné (1985) and Quellmalz (1985). It is implied that particular outcomes fit neatly into one of the three domains, but that is not the case. Outcomes may overlap two or even three of the domains. A person hitting a golf ball (psychomotor) must know what club to use (cognitive) and feels a certain way about performing the activity (affective). While the remainder of this section deals in turn with outcomes in the cognitive, affective, and psychomotor domains, it is not suggested that learning in any one of these domains is independent of learning in another. Learning generally involves all three domains simultaneously. They are organized separately here for purposes of the discussion only.

Evaluating the Cognitive Domain

The cognitive domain emphasizes the acquisition of skills that focus on knowledge, understanding and thinking. Bloom (1956) identified six major categories of learning in the cognitive domain. Learners recall previously learned material; they understand, apply and analyse this knowledge. They put parts of it together to form a new whole and judge the value of material for a given purpose. These categories are hierarchical and represent increasingly complex forms of learning. The taxonomy is also perceived as encompassing all possible learning outcomes that might result from instruction. Norris (1994) noted limitations in Bloom's taxonomy in that it does not tell what to teach when teaching for higher order thinking. It assumes a hierarchy and a degree of separation among categories of thinking that do not exist, and it uses the word knowledge in an idiosyncratic fashion. Nevertheless, the use of a taxonomy such as Bloom's reminds teachers of the need for the focus of instruction to go beyond the lower levels of learning. While all levels may not be covered for each segment of instruction, use of the taxonomy will help prevent serious omissions while reminding the teacher of the complexity of learning outcomes to be considered. In short, it encourages teachers to teach higher order thinking and discourages fixation of the teaching process at the lowest level of recall of knowledge.

Teachers should not stick slavishly to each level of a taxonomy when designing evaluation activities. In the cognitive domain, for example, the bottom level of Bloom's taxonomy can be regarded as requiring low-level cognitive activity, the three intermediate levels as requiring medium-level cognitive activity and the two highest levels as requiring high-level cognitive activity. While many different types of tasks can require students to exhibit low- and intermediate-level cognitive activity (knowledge, comprehension, application and analysis), students have to create their own responses to demonstrate the kind of cognitive activity required by the two highest levels of the taxonomy (synthesis and evaluation). To assist teachers in evaluating higher-level cognitive skills, a list of appropriate verbs, a description of the different levels of the taxonomy, and sample items are presented in Appendix A.

The gathering of evaluation data about the cognitive domain is fairly straightforward, because curriculum outcomes are expressed in terms of observable behaviour. Practically any source of evaluation data can be used to determine student performance in the cognitive domain. However, it is important for teachers to ensure evaluation considers all domains rather than being limited to the cognitive domain.

Evaluating the Affective Domain

Just as Bloom's Taxonomy (1956) provides guidance for writing outcomes that focus on all aspects of cognitive development, Krathwohl, Bloom and Masia (1964) identified five categories for classifying affective outcomes. Learners pay attention to a stimulus, respond to it, and place a value or worth upon it. Values are brought together and a value system that is internally consistent begins to develop. Behaviour which is consistent and predictable is governed by this value system.

The affective domain deals with feelings and emotions and is exemplified through students' attitudes, interests, value systems and appreciations. There is little doubt this is a more difficult area to evaluate than either the cognitive or psychomotor domains. For this reason, it is generally

inappropriate and undesirable to attach percentage values to affective outcomes. This does not, however, diminish the importance of the affective domain, since values and value systems are central to the learning and evaluation process (Eiss and Harbeck, 1969). It is important for teachers to keep the three domains separate for evaluation purposes. This is important, since it is possible for progress to be uneven across the different domains. Even though a student might have a negative attitude toward school, it is still possible, though unlikely, for that student to achieve at a high level in the cognitive domain.

One of the most effective sources of data for use in evaluating outcomes in the affective domain is observation. This method, utilizing appropriate recording devices such as rating scales, checklists anecdotal records and running records enables the teacher to focus upon and record achievement in the affective domain as indicated by behavior. The teacher must always be aware validity is a challenge here. It is easy to err when attitudes are inferred from behavior. Things are not always what they seem. The teacher should always bear this in mind and be forever vigilant about the conclusions drawn. Other sources of evaluation data for the affective domain include self-reporting devices such as an attitude scale, where students respond on a continuum ranging from favourable to unfavourable, and individual inventories where students check their preference for, or attitude toward, an idea or activity. Again, validity is a concern, since it depends upon the forthrightness of students. This is particularly the case where students are older. Student reports or term papers can be evaluated from an affective point of view if they are designed so the student is required to take a stand on a certain issue. Personal interviews and discussions can also be used to evaluate affective growth if care is taken to use indirect questions so that students cannot guess which answer the teacher would like to hear. A list of appropriate verbs and a description of the five levels of the affective taxonomy are presented in Appendix B.

Evaluating the Psychomotor Domain

The psychomotor domain focuses on outcomes that emphasize motor and manipulative skill development. Although this domain may contain outcomes which are common to most subjects, (such as writing in Language Arts, construction in Mathematics, using laboratory equipment in Science, painting or modelling in Art, and map drawing in Social Studies), the emphasis is on performance skills and most attention is given to it in subjects like Art, Music, Physical Education and Technology.

The development of a classification system for the psychomotor domain is more recent than those for the cognitive and affective domains. Simpson (1972) proposed a classification scheme containing seven major hierarchical categories. The learner uses sense organs to obtain cues, displays readiness to take a particular action, deals with a complex skill by imitation as well as trial and error, habitually responds to a performance act where movements are performed with confidence and proficiency, performs complex movements with skill, modifies movement patterns to meet special requirements or problem situations, and creates new movements to fit a particular situation or specific pattern. Observation is an effective source for evaluation data in the psychomotor domain, although some measurement may be used in subject areas such as Physical Education. Once teachers have prepared learning activities consistent with curriculum outcomes, careful observation using appropriate recording devices can be used effectively to evaluate progress toward these outcomes. A list of appropriate verbs and a description of the seven levels of psychomotor outcomes are presented in Appendix C.

Process and Product in Instruction and Evaluation

Debate about the relative importance of process and product in instruction and evaluation is quite common. The controversy implies there is a dichotomy between the two orientations, and often one's experiences reinforce the perception a gap does indeed exist. On the one hand, it is perceived that education consists of having students "stockpile" knowledge deemed to be of most worth. On the other hand, it is argued greater attention should be given to the kinds of processes most conducive to the acquisition and utilization of knowledge. An attempt will be made here to define the product and process orientations to instruction, and to examine possible implications for student evaluation.

Understanding Process and Product

Product can be thought of as the *what* of learning. *What* is it students should know? *What* is it they should be able to do? These *whats* are typically spelled out in outcomes, whether they are based on a program, a course, a unit or a lesson. In a teaching episode which is biased toward product, the ends are evaluated without reference to the means. All is well as long as students know or can do what the outcomes say they should know or be able to do. Generally speaking, students are passive in teaching situations oriented toward product.

Process refers to the procedure which students go through as they come to know, to value, and to be able to do. The emphasis in teaching shifts from a passing on of "*the what*" to the methods by which it is acquired and utilized. Students are taught to find information, to become autonomous thinkers, and to use knowledge to solve new problems and make decisions for themselves. In a process-oriented approach to instruction, students are active participants.

The differences between process- and product-oriented approaches to instruction are perhaps better understood by looking at an extreme example of each. In a social studies course, there is a body of content which students are expected to know. The teacher provides students with what is thought to be an exhaustive list of questions, along with the answer to each question. Students are required to "learn" this and reproduce answers to a sample of questions appearing on an examination given at the end of some episode. In this situation, the emphasis is solely on product. The process many students will go through is memorization. Any reasonable means of producing the product is justified, although actual learning experiences may be minimal.

A corrective reading program used in some schools many years ago was an example of a program which was heavily oriented toward process. Students were taught decoding skills and required to use rules to determine such things as whether a word had a long or a short vowel sound. Learning was supposedly demonstrated when a student could read nonsense words. Students were learning to read, but the program provided nothing to read other than nonsense words. In this case, a product (or content) was completely absent and learning was likely minimal as well.

Frost (1989) stated, "... quality programming and instruction are neither content based nor process based, but a wise and judicious mixture of both" (p.11). On the one hand, it is important students

learn to produce a product; on the other, how they learn to produce that product is equally important. The dichotomy is more the degree of emphasis given one over the other. As shown in Figure 2, a teacher at point A on the process-product continuum would view product as most important. Process would be given almost no attention. A teacher at point C would view process as being so important a knowledge base essential to full development may be lacking. A teacher at point B would have struck a balance between process and product.

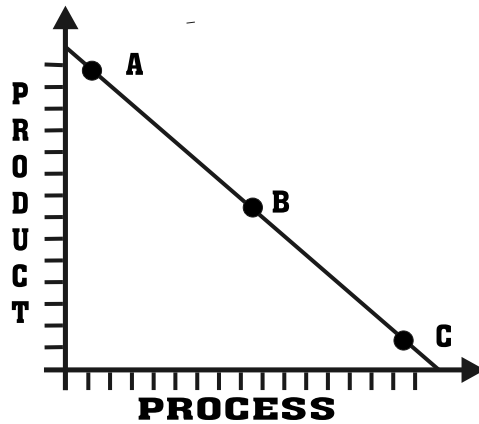


Figure 2: The Product-Process Continuum

Evaluation of Process and Product

Some questions arise regarding the evaluation of process and product. Both should be evaluated, but they cannot be evaluated independently, and the evaluation of each implies a different use of the available data. Conclusions regarding the efficacy of a process can usually be reached only through the evaluation of product. If the product is good, the process is usually acceptable. If the product is not good, the process is not good.

While the evaluation of process is usually formative, it can be summative when the process is the outcome. For example, it might be an outcome of a social studies course that students experience living in a model democratic environment. If such an environment is created, and students experience living in it, the outcome is achieved. When the process becomes the product, teachers need only observe that a student has been involved in the process. A satisfactory evaluation can be provided without any reference to the quality of any experience. The process is the product. If, however, students are to learn about such specific concepts as majority rule, there will have to be an evaluation of product to determine if students have learned what they are supposed to have learned.

Although summative evaluation is evaluation of product, this does not mean products are evaluated only at the end of a learning episode. As students go through a process, many products are generated. These products can be evaluated at any time, and these evaluations can determine how well processes are working. How successfully students use various processes can be inferred at different points in the development of a product. In addition, the extent to which students engage in various processes and the efficacy of these processes can be observed through use of various sources of evaluation data, particularly observation.

In evaluation of process, it should be realized students are active participants in a learning environment. They are doing things. In order to do things, students must have various skills. They must be able to interpret data, express themselves, participate in a group, do research and have various psychomotor skills. Such skills are usually evaluated to ensure students are acquiring those that will enable them to become active learners who learn how to find out on their own, how to learn on their own, and how to make decisions for themselves. In evaluation of product, what students know and can do is assessed in relation to intended learner outcomes.

Outcomes have to be considered in light of their implications for evaluation. What is needed is a plan that matches outcomes to appropriate sources of evaluation data. Any evaluation effort, if it is to consider both product and process orientations, has to use a wide variety of data sources. In making a decision on appropriate methods of evaluation, the emphasis on the process-product continuum should be evident from the stated outcomes.

The debate on the process-product orientations in the teaching-evaluation cycle may never be fully resolved. Although products have some intrinsic value, we should be equally concerned with the process students go through as they learn. Rather than simply learn to produce a product, students should learn in such a way they develop a large measure of independence, and learn to think and make decisions for themselves rather than depend upon someone else. Even in subjects where there is a specific product, it is still very important to give due consideration to the process students go through as they produce that product. Instructional approaches should be carefully selected to ensure product and process outcomes are achieved. It can be argued the learning process implies intellectual growth or the acquisition of new knowledge, and hence a product orientation. In that sense, product is always present, so the debate can focus only on the degree of emphasis.

Differentiated Evaluation

It is sometimes felt that unless teachers evaluate all students in a classroom in the same manner and with the same assessment sources for data, an element of unfairness enters into the evaluation. However, students have different styles and rates of learning. They must be given opportunities to demonstrate their learning in a variety of ways. Each student is unique. Each student has strengths, abilities, and areas of needs. Recognition of these strengths and needs, as well as knowledge of individual learning styles, should be the basis for adaptation of evaluation procedures. Teachers should provide for individual differences among students. Where evaluation procedures are adapted, as in the case of a student requiring Pathway 2 supports, parents or guardians as well as students when appropriate, should be involved in deciding the most appropriate adaptations. Written reports should indicate the nature of evaluation adaptations.

In the process of evaluating any student, including those with particular needs, evaluation methods should not negatively affect a student's ability to demonstrate achievement of outcomes. Just as we could not evaluate a hearing impaired child solely on the basis of an oral evaluation, we should not evaluate a child with a disability that impairs reading or writing ability solely on the basis of written tests. If teachers want to find out what students know or can do, it is essential to ensure the evaluation method itself does not impede the evaluation process. The basic tenet of this section is evaluation procedures should provide for individual differences among students. It follows, therefore, that evaluation procedures should be based on the stated outcomes but adapted according to the unique needs of the students. This is true for students doing provincially prescribed curriculum as well as students requiring Pathways 3, 4 and 5 supports as outlined by the Newfoundland Department of Education, (1998).

Consideration should be given to not deducting marks for mistakes in spelling, grammar or handwriting unless these elements are being directly evaluated. Students with difficulties in the area of spelling may misspell some words on a science test but they may be able to approximate the correct spelling well enough to demonstrate knowledge of the science material. Students may be able to use a word processor or a spell-checker for stories, essays, or projects. The intent of the assessment accommodations is not to give students on Pathways 2, 3 and 4 an advantage but to allow them to demonstrate what they know without being impeded by possible exceptionalities. There is an onus on teachers to focus on the individual needs of a student, and to use sources of evaluation data that can be used to determine clearly the extent to which outcomes of a course have been achieved.

In some cases, differentiated evaluation will consist of different means of evaluating provincial outcomes. Such differentiation might take the form of oral testing or extended time as described in Pathway 2. In other cases, students will be evaluated on outcomes of modified programs or courses which are part of Pathway 3. Generally, students requiring modified courses or programs involving deletion of, or changes to, provincial outcomes or depth of treatment of outcomes will also require many of the adjustments to evaluation as described in Pathway 2. Similarly, a student taking an alternate course should be evaluated on the outcomes of that alternate course.

In the evaluation process, the teacher is trying to determine if students have achieved established outcomes. If the outcome is to determine whether students in a Grade 3 Social Studies class understand the concept of community, methods for evaluation of this outcome must be selected. The outcome may be evaluated differently for some students. Some students may demonstrate understanding by answering written questions concerning the concept. Students with poor writing skills could use a map to explain the concepts verbally. Some students could do a research project on communities, while others could demonstrate understanding of the concept through a collage or poster. In all cases, students are showing the outcome has been met. What has changed is the source of evaluation data.

Differentiated evaluation has long been considered desirable by special education teachers and writers in the field of special education. Adaptations to the evaluation process are legitimate and educationally acceptable for children of all ages and abilities. Different sources of evaluation data should be used whenever students' needs interfere with their ability to complete a particular evaluation exercise. Adjustments to assessment must be specifically chosen to address identified needs.

What follows is a compilation of evaluation adaptations from many sources including good teaching practices. Adaptations to evaluation include adjustments to organization, adjustments to the measuring instrument and adjustments to the method of responding.

Adjustments to Organization

Adjustments to organization include flexible scheduling and setting, such as changes to setting, timing or scheduling. Changes to setting include changes to where the student takes the assessment. The assessment could be taken in a gymnasium or individually administered. Adaptive furniture could be used if desired.

Teachers should make whatever adaptations are necessary for students to be able to demonstrate achievement of intended outcomes.

Adjustments to the Measuring Instrument

Adjustments to the measuring instrument include changes to how the student accesses the test format and instructions. Examples of adjustments to the measuring instrument include but are not limited to:

- using large print or braille for a student with visual impairments;
- highlighting key words or phrases;
(Students who have perceptual problems may need clearly-typed materials with which to work.)
- using start and stop signs on answer forms;
- highlighting directions;
- providing expanded, rewritten or oral directions; (Instructions on projects or tests may need to be clarified or stated in simpler terms. Students who have difficulty following directions may need monitoring during evaluation activities to ensure they know what is expected. Often such students may be able to perform the required task but are confused by directions. Students with

short-term memory problems may need directions stated one at a time rather than three or four together.)

- providing examples;
- reading the test orally to the student, or putting the test on audio tape;
- breaking the assessment down into shorter parts, administering one part at a time, and allowing extra time if needed;
- adapting questions to a lower reading level for students whose reading ability is well below grade level;
- changing the layout and format of the test, and
- reducing, for some students, the amount of work to be done. (Students who work more slowly than others may be able to complete only three division problems instead of five. Students who would have difficulty completing a full test may be able to complete only core questions. These questions would be indicated before the test is administered.)

Adjustments to the Method of Responding

Adjustments to the method of responding are made to allow students to show what they know without being limited by a response mode the student finds cumbersome. Examples of adjustments to the method of responding include but are not limited to:

- providing more time for students who have difficulty processing or producing information;
- dividing the assessment activity into smaller parts for students who are easily distracted may need to do tests or assignments in several smaller sections and in different sessions;
- formatting answer sheets for students who have problems with organization;
- providing word banks for fill-ins;
- using portfolio assessment;
- audio- or video-taping performances for students who have trouble understanding the concept of marks, grades or comments in evaluation and encouraging self-evaluation;
- having students illustrate skills or knowledge through demonstration; (For example, students who have difficulty reading can demonstrate knowledge of safety rules by correctly operating tools. Students with weak verbal skills may be able to construct a collage or poster.)
- using oral evaluation techniques; (Students who read well but have problems with written expression can use a tape recorder to answer questions that require more than a short answer. Students with a reading disability can have questions read to them and can dictate or write the answer. Students who are strong orally can give an oral report instead of a written one.)
- making special provisions for the completion of difficult tasks; (Where a computer-scorable answer sheet is used and the student has difficulty bubbling in answers for example, the student could write answers on the test book and have them transcribed later.)
- providing a disabled student with special tools; (If the student is unable to physically manipulate a writing instrument for example, a computer could be provided.)
- modifying evaluation schemes: (In evaluating students who have exceptional ability for example, the practice of simply assigning more questions or extra work is not the best idea. Such students could be encouraged to provide several solutions to the same problem, given assignments that require more sophisticated research skills or the use of higher-order cognitive skills. Independent projects could be used to demonstrate knowledge and skills acquired outside the prescribed curriculum. Contracts could be used to specify expectations for units of work

and the evaluation scheme. Credit could be awarded where special skills and expertise are demonstrated.)

Evaluation Issues in Direct Instruction

Additional suggestions regarding adjustments to assessment and evaluation for students with identified exceptionalities were developed by the Newfoundland Department of Education (1998). In today's classroom, students are actively involved in their own learning and evaluation. For the most part, students are active learners rather than passive receivers of information. Yet, there are times when direct instruction is an effective teaching method. Such might be the case where basic skills are being developed, for example. Where direct instruction is being used, there are certain teacher behaviors which have been shown to produce the greatest learning gains in basic skill development. Many of these behaviors relate to the evaluation and monitoring functions performed by teachers. The following behaviors have been shown to be effective:

- once seat work is set, teachers should monitor students carefully, preferably by moving around the classroom (Brophy & Evertson, 1976; Emmer, Evertson & Brophy, 1979; Tikunoff, Berliner & Rist, 1975);
- seat work should always be checked (Good and Grouws, 1975);
- teachers should ensure all assigned work is completed (Brophy & Evertson, 1976);
- when teachers monitor classes through questioning, all students should be given the opportunity to participate. Although this is less critical at the intermediate and senior high levels, it is still important. Teachers should pay particular attention to having reticent students participate (Brophy & Evertson, 1976; Brophy & Good, 1986; Emmer, Evertson & Brophy, 1979; Hughes, 1973);
- unless the correctness of an answer is very obvious, teachers should always provide feedback to students on the adequacy and correctness of their responses. Feedback should be process-oriented and corrective when responses are inadequate or incorrect (Anderson, Evertson and Brophy, 1982; Brophy & Evertson, 1976; Emmer, Evertson & Brophy, 1979);
- feedback should be provided as quickly as possible (Denham & Leiberan, 1980). Students should receive grades on examinations within two or three days after they are written;
- when students are praised for specific accomplishments, they should receive it in a manner which recognizes and shows the value of their accomplishments and encourages them to compete with themselves rather than with others (Brophy, 1981), and
- even when students give less than a good effort, their dignity must be respected. No student should be ignored, belittled, harassed, shamed, or otherwise treated with disrespect (Tikunoff, Berliner & Rist, 1975).

Summary

This chapter has dealt with various aspects of the evaluation process. The first section distinguishes three types of evaluation. Pre-instructional evaluation involves the gathering of information on readiness to learn. Formative evaluation refers to ongoing evaluation during the instruction and learning process, while summative evaluation takes place at the end of a learning episode to grade, certify or select students.

Regardless of the type of evaluation used, the initial step in the process is the setting of goals or outcomes. It is not enough that teachers “know” what they plan to teach; they should clearly articulate outcomes and communicate them to students. Outcomes should guide the teaching-learning process, and become the reference point for student evaluation. They should include each domain of learning and represent lower-order and higher-order levels within each domain.

This focus on the process of evaluation has explored the process versus product approach to instruction and evaluation, the significance of differentiated evaluation, and the various findings in the effective teaching research as they pertain to student evaluation. Process and product approaches to instruction are interrelated elements, both in teaching and in evaluation. A balanced orientation gives significant treatment to both the process of learning and the actual content which students acquire. Differentiated evaluation highlights the child-focused emphasis of an effective approach to evaluation. Consideration should be given to a student’s level of physical, sensory and mental functioning, since the teacher should consider not only *what* is being evaluated, but *who* is being evaluated. As for the effective teaching research, it provides guidance for teachers where direct instruction is used as a teaching method.

SOURCES OF EVALUATION DATA

Evaluation data may be obtained from learning journals, observation, student-teacher conferences, projects, tests, performance assessments and portfolios. All of these sources, listed in alphabetical order, are discussed in this section of the document.

Useful sources of evaluation data tasks must possess the four attributes of acceptability, practicability, reliability, and validity. These attributes represent sequential steps as shown in Figure 2.

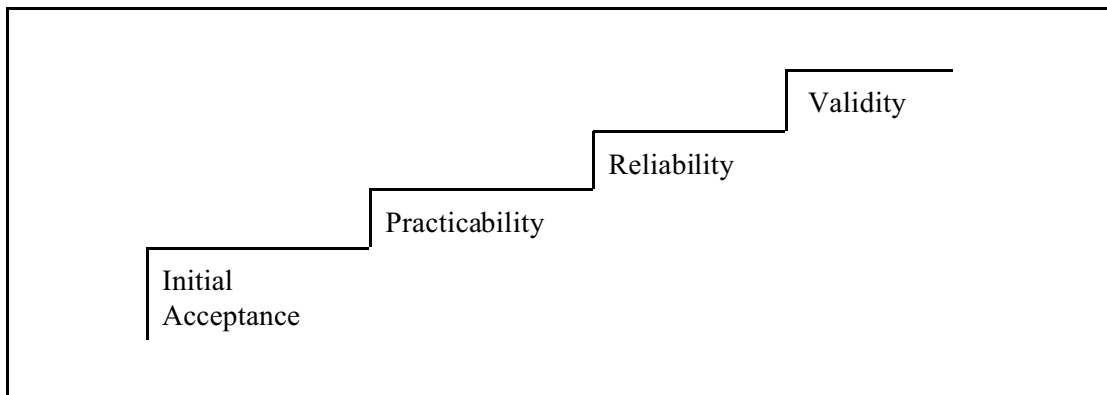


Figure 2. Attributes of Evaluation Tasks

Sources of evaluation data must be acceptable to teachers, students and parents. Unacceptable sources should be rejected and not used in evaluation activities. Acceptable sources move to the next level of the framework where practicability becomes a concern. Acceptability, therefore, is a necessary, but not sufficient, attribute of any source of evaluation data.

At the next level, practicability is of concern. If the gathering of data is onerous and time-consuming, conducting enough measurements to establish stability or reliability will be difficult and perhaps impossible. Validity cannot exist in the absence of reliability and a method of measurement that has no potential for valid use is worthless. Hence, acceptability and practicability are both necessary, but not sufficient, requirements of sources of evaluation data,

Once acceptability and practicability have been established, sources of evaluation data should be able to produce measurements that are, within the limits of measurement error, stable. If the amount of an attribute does not change, measurements of the attribute should not change. Measurements should be stable over time and across situations even where a source of evaluation data is being used in an invalid fashion. In the parlance of test theory, sources of evaluation data should provide measurements that are reliable. For an evaluation exercise or instrument to be reliable, a number of tasks drawn from the same domain in which competence is being assessed must be used (Crocker and Algina, 1986). The greater the number of tasks and the greater the homogeneity of the tasks in terms of the construct measured, the more reliable the evaluation instrument or activity. When the three attributes of acceptability, practicability and reliability have been established, validity can be investigated.

Validity is regarded as the absolutely indispensable property of any measuring instrument. Without validity, the instrument has no value. However, validity does not reside within the measuring instrument itself (Linn, 1993; Linn, Baker and Dunbar, 1991; Messick, 1989). Validity depends upon the use to which measurements are put and it lies outside, within the persons using the measurements. If measurements gathered through use of an accurate yardstick, for example, are used to draw inferences about a person's height, they will be valid. If they are used to draw inferences regarding body mass, they will not be valid. Thus, there is no such thing as *carte blanche* validity. However, measuring instruments must be able to provide measurements that can lead to appropriate inferences and decisions. That is, they must possess potential validity. Teachers can critically evaluate the validity of sources of evaluation data by determining if data can be used to answer questions that need to be answered in a teaching-learning situation.

In the section that follows, various sources of evaluation data will be considered. Each source will be evaluated according to its acceptability, practicability and threats to reliability and validity. Rarely will one source of evaluation data provide all the information a teacher will need or should have. Both reliability and validity will increase when a combination of data sources is used.

Learning Journals

Students may keep journals or diaries in which they make daily entries about what they have learned. The teacher can read the comments to determine whether or not particular outcomes are being met. Since journals may reveal how students are processing what is happening in the classroom, journal entries may sometimes highlight misconceptions which should be corrected before the unit continues. Such journals can be quite valuable in formative evaluation activities.

Acceptability, Practicability, Reliability and Validity

Learning journals are quite acceptable as a source of data concerning student achievement. They are also practical, since they are not time-consuming and they provide for increased achievement and practice in such Language Arts areas as Writing. Learning journals encourage reflective thought and metacognition as students are required to think about their learning. Since a single journal entry cannot possess reliability, the reliability of the complete data set being considered by the teacher is important. Where more than one source of evaluation data is used to draw inferences regarding achievement of outcomes, a learning journal can contribute to the reliability of the complete data set. Validity of learning journals can be established by requiring students to write about their understandings as they try to achieve the outcomes set by the teacher. Once this is accomplished, the only threat to validity would be a lack of honesty on the part of the student.

Observation

Observation helps the teacher get to know the student as an individual, because it takes into account those aspects of student performance or behavior that cannot be easily assessed through information from other sources of evaluation data. Since observation is concerned with process as well as product, it is used to monitor what a student is attempting to do, how it is being attempted, and what the student can do. For example, the process a student uses to cope with reading a difficult text will provide evidence of how that student is developing reading strategies. Observation should be widely used for evaluation in the primary school and should be an essential component of evaluation in all subsequent grades.

Observation can be formal or informal. In formal situations, tasks are prescribed and the teacher observes and evaluates. Emphasis is on what students can do and how they do it. During informal observation, the teacher does not direct the student but remains unobtrusive. This type of observation provides information on what students do on their own. Informal observation can yield much useful information, especially about affective development. When used to document student progress, results of both types of observations should be recorded. A record must be made either while the observation is being conducted or as soon as possible thereafter. Teachers should use recording devices to aid in recording the results of observations. Four common recording devices are checklists, anecdotal records, rating scales and running records.

Checklists. A checklist, as seen from the example in Appendix D, is a list of specific observable behaviours that can readily be prepared by the teacher using curriculum outcomes. Where curriculum outcomes are used to develop checklists a teacher need be concerned only with objectivity and with the adequacy of sampling to guarantee validity.

A checklist does not contain qualitative information; rather, it documents the presence or absence of particular behaviours. One big advantage of checklists is they provide immediate descriptive information. They allow for on-the-spot evaluation and, in many instances, rapid feedback about the adequacy of the performance or product. One of the most important uses of checklists is as a basis for anecdotal records.

Anecdotal records. Anecdotal records allow for the recording of indepth observations and information. They are also excellent sources of information for conferences with parents and students (see Appendix E). The anecdotal record must be based upon what actually was observed, since it is a description of some specific episode.

During observation, it is not always convenient to make detailed anecdotal records. Therefore, when the observation reveals important data about the student's development, the teacher can make notes of circumstances, dates and names. After the observation has been conducted, these notes can be used to write the actual anecdotal record. Interpretations should be kept separate from observations in the event the teacher would like a second opinion.

Rating scales. Rating scales, as shown in Appendix F, are similar to checklists, but they also include a measure of performance quality. As an evaluation instrument, a rating scale is useful in determining a student's strengths and needs and in planning a suitable program.

Running records. Wortham (1990) describes a running record as a detailed narrative of a student's behaviour that includes the sequence of events. When a teacher needs to gather data to help in evaluating particular behaviours, a running record can be established. The running record is an objective recording of everything that occurred over a particular period of time. Records may cover very short or very long periods. Analysis of the behaviour should again be kept separate from the actual record, since the teacher might like to obtain the interpretations of other professionals. The running record should be detailed enough that any reader would have no difficulty visualizing what occurred. It should contain contextual information such as the type of development observed and location as well as basic demographic information such as the age of the student. An example of a running record is shown in Appendix G.

Acceptability, Practicability, Reliability and Validity

Observation is widely accepted as a source of evaluation data. This data source requires more time and preparation than some others and might be regarded as less practical by some. For example, it is time-consuming to write anecdotal and running records. However, the rich qualitative information they provide regarding the achievement of affective, cognitive and psychomotor outcomes makes it worth the effort. It is also time-consuming to develop checklists and rating scales, although the task is made much easier where outcomes have been written and are available to teachers in electronic format as they are in this province.

As with any source of evaluation data, there are threats to the reliability and validity of observation. Reliability is threatened where sampling is inadequate. Enough observations should be conducted to provide a picture of typical behaviour. To establish generalizability, it is also important for the teacher to conduct observations under varying conditions. For example, behaviour observed in the classroom would likely not be generalizable to the playground. Validity is threatened where the observer is biased toward the individual being observed. Teachers should take care not to be affected by such factors as what others say about a particular student, by stereotypes or by intuitive feelings. Validity can be ensured if the teacher ensures the proper attribute is being observed and observation data are gathered objectively.

Performance-Based Assessment

There is presently much interest in using direct judgements and evaluations of performance rather than indirect measures of competence (Resnick and Resnick, 1992). Although these performance-based assessments are becoming increasingly popular, their usefulness is not supported by a large body of research evidence (see, for example, Guillermo and Shavelson, 1997; Herman, Klein, Heath and Wakai, 1994; Novak, Herman, and Gerhart, 1996; Shepard and Bliem, 1995).

Such assessments have been referred to as direct assessments, alternative assessments, authentic assessments and performance-based assessments (see Rogers, 1996, for example). Although many educators regard any type of constructed response to be a performance assessment, Popham (1999) suggests genuine performance assessments possess at least the features of multiple evaluative criteria, prespecified standards of performance and evaluation which is dependent upon human judgements. Others (for example, Baker, O'Neill and Linn (1993) suggest performance assessments focus on higher order or complex skills and present real world contexts. To ensure adequacy, performance assessments should be evaluated against these criteria. In addition, the teacher should be satisfied performance on a particular task can be replicated by the student on comparable tasks, and the performance evaluated is teachable, fair and scorable. A performance-based assessment with which most teachers in this province would be familiar is the provincial writing assessment.

Acceptability, Practicability, Reliability and Validity

Some are of the belief performance-based assessments will not be widely accepted because they are often viewed as lacking in rigour (see Nutall, 1992, for example). In spite of such negative attitudes, however, there is evidence in the literature that such assessments are accepted by students (Herman, Klein, Heath and Wakai, 1994; Wise, 1992) and parents (Shepard and Bleim, 1995). Because performance-based assessments are costly in terms of money and teacher time (Baker, 1998; Dietel, 1993; Herman, 1997, Nutall, 1992; Wise, 1992), practicability is of concern. However, some researchers have argued there are solutions to these problems. Shepard (1997) suggested, for example, that one way of dealing with the problem of time would be to spend less time with students in less need of assessment. There are two types of reliability which should be of concern in an evaluation situation: reliability of the instrument and reliability of scoring. Where there is but one task, reliability cannot exist, except where the data are part of a larger set. However, where responses are scored according to more than one dimension, internal consistency of these dimensions can be calculated and reliability can be expressed as a coefficient. For most teachers, reliability of scoring will be more important than a measurement of internal consistency. Reliability of scoring can be achieved where a detailed scoring guide or rubric is used. Such a guide must be accurate without being too cumbersome and a rubric must contain descriptions of different levels of achievement on each of the dimensions evaluated. Validity of performance-based assessments can be established by ensuring tasks performed by students require the kinds of behaviours specified in the intended learning outcomes.

Portfolios

In education, portfolios refer to systematic collections of student work samples (Popham, 1999). These samples are put together in such a way they tell a story (Stiggins, 1997). Portfolios tell the story of student progress over time. They allow students, parents, teachers, administrators and others to see progress in a way that is sometimes not possible when other sources of evaluation data are used.

Popham (1999) described three types of portfolios: working portfolios, celebration portfolios, and passportfolios. Working portfolios document the progress of students. Celebration portfolios showcase the accomplishments of students. In addition to a representative sample of dated samples of students' very best work, showcase portfolios usually contain a letter of introduction, a table of contents, identification of skills or knowledge demonstrated, scoring rubrics used and the student's self-evaluations. Passportfolios are used in summative evaluation to demonstrate students have met the requirements specified for a substantial chunk of work. The teacher selects the pieces to be included in a passportfolio and self-evaluations are not necessary.

Portfolios can be used to evaluate both the process and products of learning. Portfolios are comprised of products which demonstrate student learning. While products are being created, students demonstrate the processes they use as they learn.

Pieces to be entered in a portfolio are usually chosen by the student and the teacher. They are dated so they clearly show progress over time. Portfolios are usually evaluated according to criteria developed by the student and the teacher. And, portfolios usually contain self-evaluations by the student. This encourages students to become reflective about their work. They take ownership and pride in what they have accomplished. Portfolios are owned by the students.

Three or four times a semester, teachers should hold a portfolio conference with each student. At these conferences, the teacher and the student evaluate the portfolio. Students become empowered as they evaluate their own achievement, set their own goals, and become actively involved in their own learning and evaluation. Students' self-evaluation capabilities will be enhanced, enabling them to see what they need to do in all subject areas. Motivation to improve will also likely increase.

Portfolios have the potential to promote self esteem, since they can be tailored to individual students. This can allow them to demonstrate their strengths. Another benefit of using portfolios is improved communication between teacher and student as the two collaborate on learning activities, evaluation, and goal-setting.

Acceptability, Practicability, Reliability and Validity

Portfolios are widely accepted as useful sources of evaluation data. Because the gathering of student work is an ongoing process which can be integrated into the classroom routine, portfolios are somewhat practicable. However, the portfolio process will be quite time-consuming, especially for

the teacher who is a beginner at using this source of evaluation data. Many of the portfolio processes such as selecting the entries, conducting conferences with each student, teaching students to do self-evaluations, and developing an effective storage and retrieval system for the portfolios will be time-consuming. Once procedures have been established by the teacher and students, however, many of these tasks will become routine even though they will still require much time. To achieve reliability of the portfolio, teachers should ensure pieces indicate typical rather than one-shot performances. Reliability should not be a problem where students and teachers have to select from a number of pieces to demonstrate achievement of particular outcomes, especially where the portfolio is part of a larger set of evaluation data. Validity of the portfolio is assured if contents relate to achievement of intended learner outcomes specified in the curriculum.

Projects

Projects are formal assignments related to the curriculum. They may be assigned to a student or group of students, and they usually involve some type of research or development such as constructing models, preparing reports, or producing plays. While essays may share many of the characteristics associated with the project, they are considered to be a separate entity.

It is generally not appropriate to assign formal projects before Grade 4. In Grades 4, 5 and 6, projects should be completed at school with considerable teacher assistance. Beyond Grade 6, students can exercise increasing degrees of independence in completing assigned projects. However, teachers should play a major role in monitoring student progress on projects. In the remainder of this section, the discussion has been limited to formal projects and is not meant to apply to the informal projects students sometimes do.

Advantages of Projects

Where students play a role in developing projects, they are provided with an opportunity to formulate self-motivated questions and answers. This will allow for student choice in pursuing individual interests. Such pursuits might result in the demonstration of otherwise hidden talents. Where opportunities are provided to use concrete materials in illustrating their ideas, the opportunity arises for assessing skills which would not be evident with written or oral tests. Thus, projects can provide a great advantage for students who have difficulty with writing. Where students work on projects as members of a group, it allows them to demonstrate individual or group initiative. Where students present their projects to the class, they may be used as a teaching tool.

Disadvantages of Projects

Projects may be enhanced or diminished by the home environment. If a parent completes most of the project, then the student may have learned little. Yet, such a project may be given a higher grade than one which was totally completed by a student. However, if a family member has offered advice or assisted to a limited degree, then the learning experience might have been quite valuable or even enhanced by the contribution of the family member.

A considerable amount of time will be required by the teacher for organizing, grading and planning projects. Where projects contribute to a final grade, a considerable amount of time may be required by students for a small portion of the final grade. As in the case of any evaluation activity, competition may create peer comparisons which are harmful to the self-esteem of some students.

Guidelines for Assigning and Grading Projects

- Ensure students fully understand the project and have some basic methodology formulated. The teacher should be satisfied students know how to complete a project before they begin work on it.

- Teach students to use such references as encyclopedias, vertical files, CD-Rom and the Internet to obtain information.
- Ensure projects are not too difficult so students may derive personal satisfaction from doing them.
- Allow students to start projects in the classroom, especially in Grades 4 to 9.
- Provide students with a written outline of the project. This outline should include:
 - a written description of the project;
 - the intended outcomes of the project;
 - the due date (or deadlines for phases of the project);
 - evaluation criteria (including a marking scheme), and
 - suggestions concerning sources.
- Continuously monitor the progress of students. This ensures projects are manageable and realistic in scope and time. It also ensures students do not get off track.
- Provide marks or comments regarding the student's completed projects on a separate sheet of paper. Do not place any markings on the project itself. This allows the work to be displayed. However, work should not be displayed without the consent of the student, and never for comparison or competition, unless it is assigned for that purpose.
- Allow students to volunteer for group projects with the understanding each member receives the same grade for the completed work.

Evaluation Criteria for Projects

Two aspects of projects to be evaluated are process and product. Process refers to the activities which lead to the completed project, or product.

Process evaluation. Information gathered by considering the following questions can help the teacher evaluate achievement in the affective and the cognitive domains.

- Did the student indicate a willingness to be involved in the assigned task?
- Did the student undertake extra work voluntarily?
- How much personal responsibility was assumed by the student?
- Was the student open to suggestions from teacher, peers and other sources?
- During teacher-student consultations, peer discussion groups, or progress reports:
 - Was the student always aware of the purpose of the project?

- If working in a group, did the student encourage others toward completion of the project?
- Did the student remain topic-oriented throughout the project?
- If working in a group, was the student sensitive and attentive to needs of other members?
- Did the student demonstrate the thinking, research, communicative and group participation skills necessary for locating and using information?

Product evaluation. To evaluate the product, it is necessary for the teacher to compare the project to the intended outcomes. A grade assigned to any dimension of the product should indicate the extent to which it has been shown intended outcomes were achieved.

A framework for assigning a project is contained in Appendix H.

Acceptability, Practicability, Reliability and Validity

Projects are widely accepted as sources of evaluation data. They are not as practicable as other sources because of the tremendous amount of work the teacher has to put into the planning and supervising of assigned projects. Once the teacher has gone through the process once and developed project assignment procedures, they subsequently become less onerous. Since a project is only one item, reliability of this source of data cannot exist, except as part of a larger set of evaluation data. Reliability of scoring can be established where projects are scored according to scoring guides. Validity will exist where completion of a project results in demonstration of the achievement of outcomes stated in the curriculum. Validity is threatened where students complete projects outside the classroom. If the student has received help, the project might not represent that student's work. Also, where students work in groups, it might be difficult for the teacher to ensure that individual students each contributed to the completion of the project.

Self-Evaluation Reports

The self-evaluation report is an effective technique for teaching students to accept responsibility for their own learning. Self-evaluation provides an opportunity for students to examine their own progress and to analyze their own strengths and needs. When students are aware of their own needs, motivation to improve will likely increase. Self-evaluation exercises can be used at any time, and they can be combined with other sources of evaluation data. Portfolios, for example, usually contain a self-evaluation dimension. To help students think about what they are doing and what they need to do, they should be taught to ask the following questions during a self-evaluation exercise:

- What have I done well?
- What have I done not so well?
- What do I need to do now?

Acceptability, Practicability, Reliability and Validity

Many believe evaluation is the responsibility of the teacher and self-evaluation has no place in the classroom. To some extent, that is true. Summative evaluation is the responsibility of the teacher. Although self-evaluation reports can be used to help the teacher arrive at a decision, the final decision is ultimately the responsibility of the teacher. When students do self-evaluations as part of the teacher's evaluation plans, they cannot help but obtain a sense of the direction in which they need to move. Therefore, it can be a useful source of data in formative evaluation. Self-evaluation is practical, but students have to be taught to do it. Because a self-evaluation is but one item, reliability cannot exist, but reliability of the complete data set being considered by the teacher is of concern. The teacher can ensure reliability of the evaluation process by using more than one source of evaluation data to draw inferences regarding achievement of outcomes. Validity will result where students evaluate their progress toward the achievement of course outcomes. There is a threat to validity where students are not honest about their efforts and achievements. An example of a self-evaluation form is shown in Appendix I.

Teacher-Student Conferences

The teacher-student conference, like observation, is a practical and direct way for a teacher to develop a full picture of a student's growth and development. It provides an opportunity for the teacher to focus on an individual student. During an interview, it is possible for the teacher to focus upon many aspects of a student's development (Hiebert, 1991; Paris & Ayres, 1994).

Where the teacher structures the interview to answer particular questions, these questions should be written down before the conference. The interview environment should be relaxed and non-threatening. The teacher should ensure the student is at ease. The student should be encouraged to do most of the talking or ask questions. As in the case in any evaluation activity, results of conferences should be recorded.

Acceptability, Practicability, Reliability and Validity

The teacher-student conference is an acceptable source of evaluation data. Its use will result in the gathering of valid qualitative evaluation data where the student provides information regarding achievement of outcomes. It is a practical source of evaluation which requires the teacher to structure the interview and prepare a site beforehand. Because reliability cannot exist within a single self-evaluation, reliability of the complete data set being considered by the teacher is of concern. There is a threat to validity where students are not honest with the teacher, such as might be the case where the students provides responses they believe the teacher wants to hear.

Tests

A source of evaluation data widely used in the classroom, especially at the intermediate and senior high levels, is the teacher-made test. Teachers need to have a clear understanding of types of test items, test assembly, test administration, testing practices, validity, and reliability.

Types of Test Items

Item type must be carefully considered when tests are being constructed. A major decision is the relative percentage to be assigned constructed-response items (essay and completion) as opposed to selected-response items (multiple-choice, true-false, matching). Constructed-response and selected-response items are discussed below.

Constructed-response items. These items force students to develop their own answers, as in the case of essay or completion items. While such items are usually easier to construct as seen from the guidelines contained in Appendix L, the subjectivity involved in scoring, and the time required for marking impose limitations.

The general guiding principle is constructed-response items should be used to test achievement of outcomes that focus specifically on the generation of ideas. The achievement of such outcomes can be measured only through the use of such items.

As a guide for teachers in writing test items, a list of directing words corresponding to each level of Bloom's taxonomy is provided in Appendix A. The list is not meant to be exhaustive, nor are the words meant to be confined to one category. Some directing words can be used in categories other than those in which they are listed.

Good constructed-response questions are not easy to develop, but they are easier to develop than multiple-choice items. When groups are small, and more time is available for marking than for test development, the use of constructed-response questions may be quite defensible. Also, if examination questions are not to be reused for security reasons, it might be wise for teachers to develop constructed-response questions, since a constructed-response test takes less time to construct than a selected-response test. Finally, where the purpose of instruction is to develop ability in written expression, it would be more appropriate to use constructed-response questions.

Constructed-response items have some serious limitations. One of their biggest weaknesses is the subjectivity involved in the grading of responses. This subjectivity makes it difficult to produce reliability in grading, because such extraneous factors such as writing ability of students can influence grades awarded. Also, less content can be sampled, leading to problems with reliability, a prerequisite for validity. Another problem involving constructed-response items is, unless they are extremely well written, students may be unsure of exactly what they are to do. Vague directing words such as "discuss" provide the student with very little guidance. The answers given to such

questions are generally so wide-ranging teachers may have problems in grading. Marking guides could never cover all possible responses to such vague questions.

Selected-response test items. These items require students to select an answer from a set of alternatives. The major advantage of these items is scoring ease and accuracy. The most common types of selected-response items are true-false, matching, and multiple-choice. All of these item types have various strengths and weaknesses and each has a set of guidelines that should be followed closely when items are being constructed. These guidelines are provided in Appendixes M-1 through M-3.

In general, selected-response items should always be used when outcomes dealing with the lowest level of Bloom's taxonomy are being evaluated. However, their use should not be restricted to testing low-level outcomes. If properly developed, selected-response items can be used to measure intermediate-level outcomes that require students to comprehend, apply and analyze. As noted earlier, selected-response items cannot be used to measure outcomes which require students to generate ideas.

Selected-response tests require much less time for students to complete and teachers to mark, and measurement becomes more efficient. Since more items can be included on a selected-response test, and scoring is less likely to be biased, reliability is higher than on tests requiring constructed responses. In practice, teachers may often find the best tests in many subjects will contain both selected- and constructed-response items. Selected-response items should be used only to measure outcomes at the medium and low levels of Bloom's taxonomy. Constructed response items should be used only to measure outcomes at the high levels of Bloom's taxonomy.

Good selected-response items do not permit correct responses on the basis of recognition alone. They require examinees to develop a basis for choice through creative, original thought (Ebel, 1972). The argument has been made that constructing an answer is no more complex or difficult than choosing one. In fact, studies have shown scores on selected-response tests and scores by the same examinees on constructed-response tests measuring achievement of the same outcomes correlate almost perfectly. One point that should be considered is selected-response items allow more time for reflective thought on the part of examinees, since reading requires less than ten percent of the time writing requires, giving students more time to think.

Test Assembly

Early research claimed it made little difference how tests were assembled. However, in recent years, a number of researchers have found this to be incorrect. The following guidelines should help remove many of the problems associated with construction and organization of tests.

- Items should be at a reading level about two years below the grade level of students being tested.
- Items should be written according to the guidelines in Appendixes L and M of this book.
- An item should *not* be based on the response to another.
- An item should *not* provide a clue for another item.
- Items should cite authorities for statements that might be considered debatable or based on opinion.

- Items should be written in a straightforward, simple manner.
- Items should use the simplest method for requiring a correct response.
- Test directions should be clear and complete.
- Items should be grouped by type.
- Items should be grouped according to instructional content.
- Items, within groups, should be arranged in order of increasing difficulty.

Test Administration

When administering tests, teachers should give students every opportunity to do well. Thus, a careful consideration of factors that may affect test results is important.

- **Physical setting:** Every effort should be made to provide an atmosphere conducive to students' comfort. This should include ample space between students, proper lighting and ventilation, comfortable temperature, good air circulation, and freedom from distractions occurring outside the testing room. Timing might help to avoid many problems.
- **Psychological conditions:** Teachers can help reduce anxiety by adequately preparing their students. This is best accomplished by reviewing the content to be covered by the test and providing students with sample items. Tests should never be given as punishment.
- **Time:** Students should be given adequate time to write classroom tests.
- **Directions:** All students should understand clearly what they are to do.

Testing Practices

Block Testing. Testing, followed by appropriate formative activities, can be a valuable learning activity for students as well as an evaluation activity. However, the setting aside of a block of time for testing may be counterproductive, because students are disrupted and deprived of instructional time. As the research on effective teaching indicates, time on task is positively associated with learning gains in the basic skill areas. The time involved in large-scale formal testing with little follow-up activity may be uncorrelated or even negatively correlated with learning gains. With an approach to evaluation where many sources of data are used to support a professional judgement, the need for comprehensive testing would likely diminish.

Test-writing skills. Taking tests properly is a skill that should be taught. Students should be taught how to respond to test items, how to budget their time, how to approach multiple-choice items, and so on. The effect of teaching test-writing skills to students is to reduce the error variance in test scores. This will increase the reliability of the test.

Provision for success. The effects of failure can be very profound, especially if it is continuous. It is important students experience success on tests, especially at the beginning of the term. There are a number of ways of creating opportunities for success. For example, students who do not make a satisfactory grade on a test could be allowed to rewrite a parallel form. However, teachers should be careful not to create in the minds of students the attitude that success will come automatically.

Spreading students out. While measurements will result in students being spread out on a continuum of achievement, this will occur naturally. Tests need not and should not be created for the purpose of spreading students out. Time should never be used as a discriminating factor. In other words enough time should always be provided for students to write tests.

Appropriate tasks. It is unfair to require students to do something on a test that is difficult to do under the pressures of time. For example, it would not be appropriate to require students to write a poem or produce three essays on an examination.

High achievers. Some teachers feel it is unfair to high achievers if they are not given the opportunity to stand above the crowd on a test. Such an opportunity may be provided by including bonus items of greater-than-average difficulty on a test. Marks gained on these items can be used to cancel out those which were lost through careless mistakes. Thus, high achievers have the opportunity to make a higher score than would ordinarily be the case and other students are not penalized.

Surprise tests. A good evaluation system has no surprises. Surprise tests have no place in an evaluation system, because students have no chance to prepare in order to show teachers what they can do. Students should always know what is to be tested, how it is to be tested, and when it is to be tested.

Ambiguous Test Items. Tests should provide the teacher with information on what students know and can do. It is not possible for students to show teachers what they can do on a test unless they are absolutely sure of what is being asked by each item. Despite the best intentions of teachers to ensure items contain no ambiguities, students will sometimes have questions about particular items. If students indicate during a testing session they are having difficulty understanding particular items, the teacher should clear up any difficulties students might be experiencing. Teachers should never include on a test new item types with which students are not familiar.

Deduction of marks for poor English. Tests measure achievement of course outcomes and should not be used to discipline students. Often, marks are deducted when students' writing is poorly organized or standard English is not used on a test. This is not always a defensible practice. If an outcome of a course states students shall write using standard English, the teacher is justified in deducting marks if the student does not use standard English. If the outcome being tested makes no reference to such requirements, however, it is unfair to deduct marks. If marks are deducted in this case, the feedback received by students through the mark is not clear, since a mark reflects achievement on two outcomes - the one being tested and the use of standard English. How much of the mark is attributable to achievement on either of the outcomes is not clear. This does not mean teachers can ignore the quality of students' writing on tests. On the contrary, when weaknesses in writing ability or language skills show up in any work, a strong signal is sent to the teacher that there is a need to teach standard English and require its use on all written work. However, the deduction of marks is not a good teaching strategy.

Immediacy of feedback. Since many tests given throughout the school year would be used in a formative manner, the feedback they provide to students is very important. For feedback to be effective, however, it should be immediate. Although it is practically impossible to provide immediate feedback on test results, it should be provided as soon as possible after tests have been

written. If teachers find it very difficult to grade and return papers within a space of two or three days, the format of the test should be questioned. Any outcome that can be measured with the use of a selected-response item should not be tested by a constructed-response item. Selected-response items are difficult to develop, but once a teacher has a bank of them, savings in marking time will be tremendous. Teachers often argue constructed-response items in such areas as Mathematics allow them to observe the errors students are making, allowing them to take remedial action. This is certainly the case, but multiple-choice items with distracters containing common student errors can provide the same kind of information in a very efficient manner.

Provision of formulae. With respect to mathematical and scientific formulae, it is important students know where they come from, be able to apply them, and recognize where they should be used. It is usually not important that students memorize them. Therefore, such formulae should be provided whenever they are to be used, unless it is a stated outcome of the course that students should memorize them.

Structure for extended-response items. Students need structure to answer questions that require an extended response. The amount of structure may vary from specifying the approximate length to listing the major points to be discussed. Students should also be informed beforehand of the criteria according to which extended-response items are to be evaluated.

Averaging marks. When teachers have to use grades to help determine the final or term standing of students in certain subjects, average grades are often calculated and used. This may be defensible in subject areas composed of discrete units where subject material is not sequential. It is not defensible, however, to average grades in subject areas such as Mathematics, where conceptual development and skill acquisition are sequential and cumulative. To average in these subjects penalizes a student for having failed to achieve an outcome tested on an exam in October, even though the student may have achieved the outcome in November.

Using test results. Unless all the outcomes of a course, excluding the affective, can be measured through tests, it is unfair to evaluate a course solely on the basis of tests, because test results provide a limited view of student performance. For a comprehensive evaluation, where progress on non-measurable outcomes is considered, other techniques should be used to gather achievement information. However, a student who passes a final comprehensive test in a subject area that is amenable to measurement should pass the course.

Using publishers' tests. Publishers' tests often do not measure achievement of the outcomes of a program as they have been defined in curriculum documents. They do not always do a good job of measuring the outcomes stated in the textbooks they accompany. Only those that measure the outcomes outlined in curriculum guides and have been constructed in accordance with item construction techniques outlined in this book should be used in the classroom.

Surface quality of tests. All test scores are composed of a true score plus or minus error score. One possible source of error in test scores is surface quality of the test. If students have trouble reading a test that is poorly written or poorly copied, the amount of error in their scores will be larger than would normally be the case. Teachers should ensure tests are very clearly written and very clearly copied.

Reliability

One conception of reliability is concerned with the amount of random fluctuation or error there would be in individual test scores if a test were given over and over. Another conception of reliability relates to internal consistency. In order for the items or tasks on a test to be internally consistent, they must measure tasks from the same domain. When items are judged to be valid measures of the outcomes that comprise a well-defined domain, reliability or internal consistency will follow naturally provided tests contain enough items. A test can be reliable without being valid. An inaccurate or invalid scale, for example, can provide consistent or reliable measurements of mass, even though these measurements are not valid.

There are many ways of increasing reliability, but these are restricted to the domain of norm-referenced testing. As with any source of evaluation data, the main consideration for teachers with respect to reliability of tests is that they contain enough items. There should be five or six selected-response items testing each outcome, for example. If true-false items are used to test the same outcome, there should be more. In a test containing constructed-response items, reliability is enhanced when there are clear guidelines for the marking of answers.

Validity

For a test to be valid, it must be an accurate reflection of the domain it is meant to test. A domain can be conceptualized as an infinite number of tasks. Each task has two dimensions: content area and cognitive processes. A test can be conceptualized as a collection of tasks selected from the domain in such a way that content areas and cognitive processes of the test accurately reflect the content areas and cognitive processes of the complete domain, whether that domain be a unit or a complete course. Two things need to be considered when the validity of a test is established.

First, a sampling plan must be developed to sample the domain scientifically. This plan is referred to as a table of specifications, and it must contain two dimensions. The vertical dimension of the table contains the topics of instruction. The horizontal dimension is comprised of the cognitive processes in which students are expected to engage. The body of the table is comprised of numbers which provide weights or rankings of importance of both dimensions of the domain. When a domain is sampled scientifically through use of a table of specifications, one is able to generalize beyond the items that comprise the test to a complete domain or some portion of it. In other words, a percentage grade on a test indicates more than proficiency on that test. It indicates proficiency on the complete domain or the portion of the domain being sampled. A sample table of specifications is shown in Appendix J. The table of specifications can be used to develop a detailed plan of the examination. Such a plan, developed from the table in Appendix J is shown in Appendix K.

The second requirement of validity is that each item measure the outcome it is meant to measure. This can be established through logical analysis where a teacher examines an outcome and the item used to measure that outcome (Roid and Haladyna, 1982). Based on this analysis, the teacher judges whether the item is a measure of the outcome. If the item is judged to measure the outcome it is meant to measure, it is valid. If not, it is invalid. Empirical procedures requiring the ratings of judges can be used, but they are not necessary.

Acceptability, Practicability, Reliability and Validity

Tests are quite acceptable to measure achievement of educational outcomes. However, it is recognized today tests are more useful at some grade levels than others. It is also recognized test results alone rarely provide a comprehensive picture of achievement. To provide a complete picture, test results should be used in conjunction with evaluation data gathered through the use of other sources.

Tests are practicable, but care must be taken to sample the domain correctly and to create rigorous items according to the guidelines provided in Appendixes L through M-3.

Care must be taken to ensure reliability and validity if the test is to have value as an evaluation tool.

Work Sample Files

Work sample files containing dated work samples a student has produced throughout the year provide useful evaluation data. If the samples are systematically collected from all areas of study regularly, the work sample file can provide direct evidence of a student's overall progress.

The work sample file can be used for both formative and summative evaluation. It is a useful aid in talking to parents about a student's strengths and special needs. Work sample files are also useful for providing students with a number of products from which they can select pieces for inclusion in a portfolio.

Acceptability, Practicability, Reliability and Validity

Since they provide direct evidence of students' achievements, work sample files are acceptable for demonstrating student progress. Work sample files are practicable, since it is not time-consuming to collect the samples, and it does not require a lot of effort. Reliability can be ensured where work samples are chosen such that they indicate typical student behaviors and are replicable. Validity can be ensured if the work samples indicate achievement of intended learner outcomes which are specified in the curriculum.

Summary

In this chapter, popular sources of evaluation data for use in pre-instructional, formative and summative evaluation have been considered. While different sources may be more appropriate at different grade levels, all teachers should use a variety of sources as they evaluate student achievement. They should choose those sources which best provide the desired data. Where more than one source of evaluation data are used, validity will be enhanced. All teachers should be aware of the threats to the reliability and validity of each source.

INTERPRETING AND USING EVALUATION DATA

When a student's progress is being summarized for reporting or for making an educational decision, all relevant information should enter into the teacher's judgement. Although little has been said about teacher judgement in this handbook, it is an important ingredient in any evaluation decision. Teacher judgement is necessary for evaluation to have occurred. Informed teacher judgement is necessary in any evaluative exercise, and nothing should impinge upon the right and the responsibility of the teacher to make that judgement. For the most part, evaluation is based on objective data, but teachers may have to go beyond objective data gathered in the classroom. Where a student has missed many summative evaluation activities, the teacher may have to make a professional judgement, in the absence of hard data, regarding the student's final standing. When such factors as a traumatic event in the family is judged to have had an influence on student achievement, they should be taken into account. It is a responsibility of all teachers to judge evaluation data and make defensible decisions about grading, reporting and promotion.

Grading

In educational debates, many issues have arisen regarding grades. On the negative side, grades may serve to divide teachers, negatively affect the students' self concepts, lack meaning for students or lack reliability. Grades can also become ends in themselves for students and teachers, and become more important than learning.

In spite of all the weaknesses in the application of grading in the classroom, grading is a necessary and pervasive practice in education. And, there are positive benefits of grading. Grades motivate students. They can be used to provide a basis for decisions about promotion, graduation and subsequent schooling. They can be used to recommend students and graduates for employment and to award special honors. They provide information for counselling and informing parents and the students themselves of the progress being made toward the achievement of intended learner outcomes.

Types of Grading Systems

The most pervasive systems used for reporting achievement are percentages, letter grades, or pass/fail. In recent years, however, rubrics have become popular in grading student responses. Percentage grades and letter grades share many of the same advantages and disadvantages. Both are convenient to use for administrative purposes. Both are readily accepted by post-secondary institutions, since they are fairly easy to interpret and they are good predictors of future grades. For some students, percentage and letter grades may be motivators. One advantage percentage grades have over letter grades is they are simple and quantitative, and they can be entered as they are into many statistical calculations. This is difficult to do with letter grades, since they usually represent a range of achievement rather than a precise quantity. Where numbers are desired, the middle of the range can be used as a substitute for a letter.

Letter and percentage grades also share many of the same disadvantages. They do not give a detailed indication of students' strengths and needs, and they are sometimes based on more than one variable. This makes them difficult or impossible to interpret, and reduces validity and reliability. The use of percentage or letter grades contributes to anxiety and excessive competition among students, and may inhibit the development of self-confidence. Grades given in different schools by different teachers may not be comparable, since different standards are often applied in arriving at them. A disadvantage which does not apply as much to letter grades as it does to percentage grades relates to interpretation. Percentage grades are often interpreted as if they were perfectly valid and reliable and without measurement error.

Rubric scores and letter grades represent a kind of data different from percentages and letters. When a rubric or letter grade is used, student work is classified according to a qualitative classification scheme in the case of a rubric or a quantitative classification scheme in the case of letter grades. Letter grades borrow heavily from the continuous percentage scale, and the only difference is letter

grades are not truly continuous, since they take on only certain points or ranges of the continuous scale. An A might have a value of 85 or 95, or it might represent a range of achievement of 85 - 95. The numbers associated with letter grades do have the properties of numbers, however. They can be manipulated through the use of arithmetic processes. Also, one can convert letter grades to numbers or numbers to letter grades where the need arises.

Rubric scores are a different matter. Usually, a rubric score is in the form of a number from 1 to 5. Numbers associated with rubrics, however, have no properties of numbers. They are used for naming and they can be ordered, but spaces between numbers are not equal, and there is no zero. More importantly, their descriptions are purely qualitative. Since student products are categorized according to qualitative descriptions, the data resulting from use of a rubric are categorical rather than continuous. Hence, it makes no sense to convert rubric scores (qualitative and categorical) to percentages (quantitative and continuous). If a student's performance is classified as three on a five-point rubric, it does not mean the student scored three out of five, or sixty percent. It means the product matches the description contained in level three of the rubric. The description will focus on the characteristics of the performance, and by comparing it with other descriptions, the teacher, the student, and the parent can see why the performance is not a 4 or a 5 (or a two). In spite of this, there are some who desire to convert rubric scores into percentages. As Glass and Stanley (1970) have said in a highly-regarded authoritative text,

There is no law preventing one from adding, subtracting, multiplying, etc. numbers that have been assigned to objects by ordinal measurement. However, the results of these operations may reflect nothing about the amounts of the property in question that the objects corresponding to the numbers possess (p. 10).

The official position of the Division of Evaluation, Testing and Certification is rubric scores should not be converted to percentages. If percentage grades are desired, student performances should be graded twice - categorically in comparison to a scoring rubric and continuously with a grade which quantifies the extent the performance reflects the attributes of a model response.

No letter or percentage grades are used in a pass/fail grading system. Pressures of competing for percentage or number grades are removed, and creativity and exploration of new subject matter are encouraged. This system promotes an atmosphere conducive to learning, although some pressure may be felt by students close to failing. The amount of information provided by pass/fail grades, however, is less than that provided by percentages and letters. Such grades do not discriminate among students of differing abilities, and they are poor predictors of future academic success. Perhaps the most severe limitation of pass/fail grading systems is they encourage minimum rather than maximum achievement.

Bases for Grading

Grades can be based on such variables as progress made over some period of time, extent to which the student has achieved curriculum outcomes (criterion-referenced), achievement in relation to ability (self-referenced) and achievement in relation to others in the class (norm-referenced). Additionally, some teachers use such affective attributes as effort, interest, cooperation, attendance,

neatness and diligence as bases for grading, but these are seldom used as complete measures which are independent measures of achievement of outcomes.

A grading system which uses more than one variable for grading is unacceptable, since grades lose their meaning when they are based upon a number of variables. Oosterhof (1994) noted that including multiple variables or traits in a single grade substantially reduces their reliability and validity. According to Oosterhof, the reliability of grades can be improved if they are based on measures of a single variable, measure current status as opposed to change and improvement, are based on more as opposed to fewer measures and are based on similar criteria when assigned by different teachers.

If more than one variable is to be used for determining a grade in a subject, a general rule is a separate grade should be reported for each variable. If grades are to be based on interest, effort, progress and achievement, there should be a separate grade for each of these variables. If this tends to make grading too complicated, it is better to omit some grades rather than lump them all together.

When a student's achievement is summarized, relevant information from all data sources should be used. Summative evaluation should be a general assessment of a student's performance and should not be seen as an opportunity to average the grades which the student obtained from various evaluations. When grades are reported to parents or guardians in term reports, teachers may wish to give a general overall rating using a limited number of categories. For example, at primary, elementary and intermediate levels, a four-point grading scale such as the following could be used:

- 4 Meeting all the outcomes of the program.
- 3 Meeting most of the outcomes of the program.
- 2 Meeting some of the outcomes of the program.
- 1 Meeting few of the outcomes of the program.

Such a grading scale could be supplemented by any kind of data which adds to the information provided by the report. Such information might take the form of anecdotal comments that identify outcomes with which a student is experiencing difficulty, or percentage grades where appropriate.

Including vague statements such as "can do better" or "should try harder" on a report should be avoided, since they are essentially meaningless. Where a student is experiencing difficulty, very specific information should be provided.

Reporting

Reporting refers to the provision of information regarding the achievement of students. Information may be quantitative or qualitative or some combination of both. Quantitative information usually takes the form of numbers such as marks, while qualitative information is usually in the form of a descriptive summary of student achievement in relation to specified outcomes. Reports should be neither completely quantitative nor completely qualitative. Instead, there should be a balance between quantitative and qualitative information on a report card.

While numbers will inevitably comprise part of reports at the higher grade levels, numbers often hide more information than they provide. In addition to providing numbers, teachers should give a balanced account of students' strengths and needs, using whatever information has been acquired through use of the full range of evaluation data sources. Much information, instead of resulting from the manipulation of numbers, will be the result of the teacher's use of professional judgement. Where it would clarify information contained in the report cards, student work samples could accompany report cards.

Outcomes provide a framework for teaching, learning, evaluation, and reporting. Because they fulfill such important functions, emphasis should be placed upon outcomes from the very beginning of instruction. They should be communicated, in student language, to students and parents at the beginning of an instructional sequence such as a chapter or unit.

The report card should paint a picture of the whole child. It should not be restricted to growth in the cognitive domain but should also speak to the psychomotor and affective development of students. The report must contain honest information, but at the same time, the sensitivity of the child and the importance of the student's positive self-concept should be respected. Comprehensiveness of the report card is enhanced when a report is linked to other report cards issued earlier in the year. For the sake of comprehensiveness and continuity, teachers should make conscious attempts to include such linkages in their reports.

The usefulness of student self-evaluation activities is well established in formative evaluation. It has also been shown students achieve best when they are actively involved in their own learning. Students can play a role in creating their own report cards. Where students have taken an active role in their own evaluation, the report card should be signed by both the teacher and the student as "co-evaluators." Students can also be involved in reporting by being part of the post-report conference with parents and teachers.

Quality of information communicated in report cards will depend upon a number of factors. Good information could be rendered useless if the report card is difficult to interpret. Whether or not users of report cards are familiar with particular formats or whether they desire such formats will affect the quality of information. Direction of communication flow will influence the value of the information communicated. Communication that flows in one direction will not be as rich as communication that flows in more than one direction. And, while comparisons can be odious,

students are always referenced or compared to some kind of standard in report cards. Some comparisons can be meaningful but others can be meaningless in terms of the growth of the child. Likewise, the attributes reported upon must be meaningful in terms of the growth and development of the child. Finally, quality of communication can be affected by its continuity. Even where information has the potential to be of high quality, its value will likely diminish if its transmission is sporadic. In the remainder of this section, a review of the literature regarding the views of various stakeholders on these issues is presented.

Interpretability. Interpretive aids cannot help but make communication on report cards clearer, and they should be used wherever possible (Chansky, 1975; Waltman and Frisbee, 1994). Even those aids will be of little use where different teachers use different grading standards. There is a need for the symbols used by teachers, at least at the school if not the district level, to have the same meaning (Wiggins, 1996). If symbols are explained and standardized across districts, this would result in clearer communication and make the task of explaining report cards to parents easier. A step toward removing variation in grading standards among teachers can perhaps be best accomplished through the use of scoring rubrics and exemplars. This should not be difficult where curriculum is centralized as it is in this province.

Format. Users of report cards, particularly parents, prefer for report cards to contain a mixture of grades, narratives, checklists and the like rather than one type of information (Bursuck, Polloway, Plante, Epstein, Madhavi and McConegy, 1996; Hall, 1990; Hoffman, 1995; Olhausen, Powell and Reitz, 1994; Scott, 1992). This is desirable, since different kinds of grades would likely draw on information from a variety of sources. Resulting grades are therefore likely to be more valid and reliable than those based on information from one source. It is also desirable that report cards contain narrative information. Narrative reports convey qualitative information which supplement the quantitative information provided by grades. The report card would likely be more valid if it contained both kinds of information, since more information could be provided regarding achievement of educational outcomes.

In addition to narrative information, report cards often contain comments. Use of a computer greatly simplifies the task of providing comments on a student's report card. However, use of a computer in the preparation of report cards can impersonalize what should be a very personal process, particularly if teachers select from lists of comments prepared by software publishers. If teachers use a computer to help write such comments, impersonalization can be avoided if they use software to write their own comments rather than select pre-written or "canned" comments.

The validity of the report card is enhanced when the report card contains predictions regarding the final standing of students in particular courses. At the same time, users of report cards must be made aware of the assumptions that underlie such predictions and that predictions can be wrong whenever any of the myriad of factors which affect student achievement changes. For example, one such assumption in any prediction is a particular student will continue at the same rate. That may or may not be the case. Teachers should be aware predictions of failure are particularly dangerous, because such predictions can become self-fulfilling.

Direction of Communication Flow. For the most part, report cards represent one-way communication. Although some report cards contain spaces for parents to respond (Pawlas, 1985; Perkins and Buchanan, 1983), most have no such provision. Not only is communication one-way, but report cards are usually designed exclusively by teachers. Some researchers suggest communication would improve if all stakeholders were involved in designing report cards (Frisbie and Waltman, 1992). Communication improves to two-way where parent-teacher conferences are held. This could be improved to three-way communication where students are involved in those conferences.

In many high schools, subject-teaching is common. Since teachers often interact with large numbers of students, the student-teacher relationship becomes somewhat impersonal, and no one teacher is personally responsible for monitoring student progress in all subjects or in the affective domain. Often, no one is responsible for identifying signs a student is experiencing difficulty and for initiating appropriate corrective measures. At the higher grade levels where homeroom teachers are assigned to particular classes, these teachers are usually assigned administrative tasks such as keeping attendance records. In such cases, there is no flow of information regarding the overall progress of students. In intermediate and high schools, the performance of each student in all three domains should be monitored frequently by an assigned teacher who should assume primary responsibility for reporting to parents or guardians. In elementary schools, homeroom teachers are usually given the responsibility of fulfilling this monitoring and reporting function.

The improvement of communication is desirable. Schools should do whatever they can to provide accurate and adequate information. They should endeavor to make communication two- or multi-way. In addition to conducting parent-teacher-student conferences, teachers should allow for parents and students to respond to report cards, allow a copy of the report card to remain at home, and involve parents and students in the design of report cards.

Standards of Comparison. Students' grades can be referenced to different standards. They can be referenced to the requirements of particular courses (criterion-referenced), to other students (norm-referenced), or to ability (self-referenced). Students should be graded in comparison to each other where the purpose is to select particular students for particular reasons, such as to receive scholarships. Where selections are not being made, students should be graded in comparison to the requirements of particular courses. In the research literature, there is most support for reporting that is criterion-referenced and least support for norm-referenced reporting (See, for example, Anderson and Bachor, 1993; Anderson and the Amerman PTA, 1993; Guskey, 1996). Regardless of the appropriate standard of comparison, students who do not give appropriate effort should also be evaluated with reference to ability.

In a criterion-referenced situation, student progress should always be referenced to the outcomes of instruction. Content validity is concerned with whether or not measures relate to the outcomes of instruction. This kind of validity is of great concern to the classroom teacher. If report cards are referenced to the outcomes of instruction, evaluation will likely be related to those same outcomes. Consequently, evaluation and report cards should be high in validity where student progress is criterion-referenced. Teachers sometimes reference their reporting to the amount of progress made.

Such reporting could be misleading and should not be used, since a student who has made much progress might receive high grades but be nowhere near grade-level expectations.

Attributes Reported. The research literature indicates parents are most interested in the academic growth of their children (Anderson and Bachor, 1993; Lotz and Sukorsky, 1989; Stephens, 1980). But, they are interested in their affective development as well (Chansky, 1975; Friedman and Frisbie, 1995). Since education is concerned with the development of the total person, evaluation should be concerned with affective and psychomotor development as well as cognitive development of students.

At any level, attendance is a prerequisite for achievement. Because it is so closely related to achievement, teachers should report upon student attendance. At the senior high level, attendance should be monitored and reported at the class level.

Frequency. Quarterly reports from the school to the home seems to be the norm (Lake and Kafka, 1996). Nevertheless, researchers have identified many exceptions (Lovelady, 1991; Olhausen et al, 1994; Stephens, 1980). Parents often desire more frequent reporting. There is nothing magical about quarterly reports. And, there is no compelling reason, other than tradition, to issue report cards quarterly.

Reports seem to be issued more often at the lower levels of schooling, and again, there seems to be no compelling reason for this. Information is as useful to the parents of senior high students as it is to the parents of primary students. The more information that flows between the home and the school, the better. And, there is no requirement the information flow be formal. Much valuable information can be exchanged on an informal basis through telephone calls, notes and the like. While much communication coming from the school would be one-way, providing students and parents with opportunity to respond would likely result in two- and three-way communication.

Promotion

When all evaluations for the school year have been completed, teachers and principals often find themselves having to decide if certain students should be promoted or retained. For decades, it has been known that most children who are held back a year in school do not benefit academically and do not grow as much as like-functioning children who are promoted (Dobbs & Neville, 1967; Kundert, May, and Brebt, 1995; Liaw. & Meisels, 1993; Shaeffer, and Hook, 1993). Where academic gains have been found by researchers, these gains have not occurred across all grade levels. Furthermore, gains in academic achievement have been accompanied by decreases in measures of self-concept (see, for example, Pomplun, 1988). The findings of other researchers indicate gains in academic achievement which result as a consequence of grade retention are spurious and do not last long (Dennebaum and Kulberg, 1994). Even though many studies have shown grade retention and delayed entry to be ineffective interventions, there is a gap between research findings and practice, since grade retention in the school system is common (Tanner, 1993).

Despite the fact grade retention has been shown to be an ineffective remedy for unsatisfactory achievement, there may be some students for whom retention would be valuable. The challenge lies in identifying the appropriate student at the appropriate time. The question to be answered when deciding upon promotion or retention is not whether the student has advanced far enough academically or socially, but how well the student's needs will be met as a result of the decision. Where it is not clear whether a student's best interests would better be served by retention or promotion, the decision should not be made by one person. The interests of the student should be looked at from more than one point of view. The final decision should be made by a panel comprised of individuals who have worthwhile information or knowledge to bring to the process. At the very least, the panel should be comprised of the principal, other teachers, parents and the student. It is useful to include parents or guardians, since they not only bring extra information to the process, but their involvement helps prevent the development of negative attitudes on the part of the student.

Very few students should be retained. Diagnostic and remedial help can be given by the regular teacher in cooperation with, and guidance from, the special education teacher(s) in the school. In a spiral curriculum, the previous years work is reviewed before new concepts are developed. This constant repetition and review would seem to obviate the need for student retention in all but the most extreme cases.

SAMPLE POLICY

This chapter contains sample policy statements which are based on positions outlined in this handbook. Each section begins with a broad policy statement. This statement is followed by a rationale which is a synopsis of the section upon which it is based and general guidelines which support the policy.

The handbook and policy statements could be used as the basis for developing a district policy. At the district level, there may be a desire to tailor the handbook to meet local needs. Certain issues may be expanded, or there may be a need to address some issues which have not been raised. In any event, teachers and principals should be involved in the process of making this or any other policy document fit the needs of a particular district. More effective implementation will likely be the result of such involvement.

Evaluation of Student Achievement

Policy Statement: *Student evaluation practices will be based on a philosophy of education which respects the uniqueness of each child and be conducted according to currently modern educational theory and practice.*

Rationale

Evaluation is a professional judgement of the extent to which students have achieved intended learning outcomes. It is not a discrete activity but a continuous and comprehensive process, the results of which always affect classroom proceedings.

Outcomes represent goals for students and teachers, and they provide the framework for student evaluation. These outcomes are concerned with skills and values as well as with knowledge. They are not always measurable, or even observable; they do not always specify a product, and their value may be intrinsic.

Data which can be used to evaluate students can be gathered from many sources. No data are infallible, and the suitability of a particular source often depends on the nature of the subject or discipline in which the evaluation occurs or the outcome being evaluated. When choosing a data source, primary considerations should always be the acceptability, practicability, reliability and validity of that source.

Parents are important partners in education. They deserve to be fully aware of the progress their children are making at school. Not only should parents be made aware of their children's progress, but their support for the efforts of the school should be encouraged.

Finally, evaluation must always be a purposeful activity.

Types of Evaluation

Policy Statement: Evaluation will consist of pre-instructional, formative and summative activities.

Rationale

Comprehensive evaluation requires an awareness of the strengths and needs students bring to the classroom. Where information is needed before instruction begins, it can be gathered from such sources as cumulative records, teacher-teacher conferences, student-teacher conferences and parent-teacher conferences. Teachers should be careful to remain objective and not allow pre-instructional evaluation to affect their expectations for student achievement levels.

The most valuable type of evaluation occurs during the instructional process. Formative evaluation focuses upon the process as well as the products of learning. Its main function is to document progress and identify impediments to learning such that they can be corrected or remediated as soon as possible.

Summative evaluation is used to grade or certify students after instruction has taken place. When teachers evaluate summatively, they should use all the evaluation data at their disposal. Too much emphasis should not be placed upon results gathered through use of one particular source. When data from a number of sources is used, a comprehensive picture of student achievement and progress will emerge.

Guidelines

1. Use available sources to determine the strengths and weaknesses of students.
2. Emphasize formative evaluation during instruction.
3. Gather information for formative evaluation by using a variety of data sources.
4. Communicate formative evaluation data promptly to student, parents, and the student's homeroom (or monitoring) teacher.
5. Summative evaluation decisions should be based on data from a variety of sources.

Outcomes

Policy Statement Evaluation references performance to stated outcomes from the affective, cognitive and psychomotor domains. Outcomes should be clearly stated and communicated to students.

Rationale

Essential graduation learnings are achieved largely through curriculum. General curriculum outcomes are achieved in each discipline over a student's school life. Key stage outcomes are completed by various levels such as primary, elementary, intermediate and senior high. Curriculum outcomes are achieved through course outcomes, and course outcomes are achieved through unit and lesson outcomes.

Outcomes guide teaching, learning and evaluation. If outcomes are to guide learning, students must be aware of them. Outcomes should specify what students will be able to do following instruction, and they should be in a form that is understandable by students and parents.

Education is concerned with much more than student recall of knowledge. As students mature, outcomes should be written to ensure students engage in complex cognitive processes where appropriate. Once they have comprehended knowledge, they should be required to manipulate it by applying it to solve problems, analyzing it, using different parts to form a new whole, and evaluating it.

The affective development of students is of direct concern for educators. Evaluating the affective domain, however, is difficult. Specific outcomes do not play the same role, because the provision of such outcomes cues students to act in a particular manner. The observed indicators, given the presence of such cues, would not necessarily mean an outcome has been achieved. Nevertheless, a clear idea of the affective outcome to be achieved and specific indicators of their achievement is needed. The most valuable source of evaluation data in the affective domain is observation. The results of observations can be recorded with the use of checklists, anecdotal records, rating scales and running records. Affective outcomes should focus on the student's willingness to pay attention, to take part, to attach value to things, and to begin to build a value system which gradually becomes pervasive, consistent and predictable.

It is important numerical values not be assigned to signify degree of affective development. It is important as well that progress, or lack of it, not affect progress in the other domains. Finally, it is important parents be given information on the affective development of their children. As the child's first teachers, parents are in a position to influence positively the affective development of their children and play a role complementary to that of the school.

The psychomotor domain is concerned with the motor development of students: learning to print or write, using laboratory equipment in science, baking a cake in home economics, bouncing a ball in physical education, and so on. Outcomes in this domain should be concerned with the use of the senses, readiness to act, the early stages of learning a new movement, proficient movement, skillful performance, adaption of physical skills, and creation of new movements. Observation with the use of rating scales, checklists, and anecdotal records can be used to evaluate psychomotor development. In addition, measurements can be used in some subject areas.

Guidelines

1. State outcomes in language students can understand.
2. Provide outcomes to students at the beginning of a unit.
3. Evaluate students in relation to the stated outcomes.
4. Keep achievement in each domain independent of achievement in the other domains.
5. Where appropriate, write outcomes that require different cognitive processes.
6. Use a variety of data sources to evaluate the cognitive domain.
7. Compile affective outcomes collaboratively at the school level.
8. Evaluate the achievement of affective outcomes through observation.
9. Do not use achievement in the affective domain as a basis for promotion.
10. Provide outcomes for the psychomotor domain for those courses where psychomotor development is important.
11. Evaluate achievement of psychomotor outcomes through observation. Measurements may be used where they are appropriate such as in physical education.

Process and Product in Instruction and Evaluation

Policy Statement: Process and product outcomes will be evaluated.

Rationale

Product refers to a body of information the student interacts with and is expected to know. Knowledge is transmitted to students who, at some point, must demonstrate mastery of it. In this product approach to education, the ends are all-important, and any means which lead to them are justified. Where there is a product orientation, students often interact passively with content.

In a process approach to education, the emphasis shifts from passing on information to the methods by which it is acquired. Students are more active as they learn how to find information, how to become autonomous thinkers, how to use what they know to solve new problems, and how to make decisions for themselves. Some advantages of a process approach to instruction are it increases motivation and intellectual potency, aids memory, teaches students to discover for themselves, increases self-confidence, and improves attitudes toward learning.

Orientations to teaching are not pure process or pure product. There is some concern with both. However, when product becomes an end in itself, the balance between process and product is upset, and process is a slighted partner. Likewise, in a situation where process is all-important, the product essential to the students' full development is ignored. Ideally, the curriculum should be regarded as being worthwhile and of intrinsic value, but the importance of the processes instrumental to the achievement of the product outcomes should not be overlooked.

When a balance has been struck between product and process, the extent to which students know and comprehend things, and the extent to which they can do such things as think autonomously, use prior knowledge to solve new problems and to make decisions, should be evaluated. Also, the evaluation should consider development in the various cognitive skills students need to be able to do these things.

The full range of data sources may have to be employed to provide meaningful evaluations in all areas.

Guidelines

1. During evaluation, teachers should reflect the relative emphasis given to product and process during their teaching.
2. Use different data sources in the evaluation of process.
3. Provide students with general evaluation plans at the beginning of a course.

Differentiated Evaluation

Policy Statement: ***Differentiated evaluation will be employed to accommodate students with special needs.***

Rationale

In addition to a set of skills, education should result in the development of a sense of personal dignity and worth in every student.

In the process of evaluating students, the source of evaluation data used should not negatively affect ability to demonstrate achievement of outcomes. Outcomes may have to be evaluated differently for some students if they are to demonstrate their level of achievement accurately. A student may be unable to demonstrate acquired knowledge because of some special need. However, this does not always indicate a lack of learning or understanding. A different mode of evaluation may produce completely different results. There is an obligation to focus on the individual needs of a student and to employ evaluation methods that clearly determine the extent to which the outcomes of a course have been achieved.

Each student is unique. The range of variation defies description. The school should provide for individual differences among students. Specified outcomes and sources of evaluation data may have to be adapted to meet the needs of students. Where adaptations are made, parents and students should be informed in a timely, sensitive and meaningful fashion.

Guidelines

1. Reflect the outcomes of instruction in evaluation methods.
2. Ensure data sources are congruent with the needs of students.
3. Explain evaluation adaptations to students and parents prior to instruction.
4. Involve teachers in all decisions regarding differentiated evaluation.

Sources of Evaluation Data

Policy Statement: A variety of sources will be used in the gathering of evaluation data.

Rationale

A single data source gives a limited view of student achievement. A comprehensive picture of where students stand in relation to course outcomes can be obtained only through the use of a variety of data sources.

Observation is a valuable source of data for evaluating process as well as product. To be effective, it must be conducted purposefully, regularly and systematically. Appropriate tools such as checklists, anecdotal records, rating scales and running records should be used to record the results of observation activities. The use of such sources as the portfolio, work sample folder, self-evaluation activities, interviews and learning journals can also be used to convey timely, relevant information to parents or guardians and students. Performance-based assessments can be used to provide direct evidence of the achievement of outcomes.

Projects are often assigned to students. Generally speaking, however, the younger the student, the more teacher guidance and assistance required. Projects assigned at the primary and elementary levels should generally be completed at school, and should be regarded as learning activities, not evaluation activities. At the high school level, the results of assigned projects may be used in evaluating students. Nevertheless, even very mature students need guidance when projects are assigned. Students should be provided with the intended outcomes of the project, a description of the task, relevant dates, references and evaluation criteria.

Tests should be designed to measure student achievement of measurable outcomes. Students should always know what is to be tested, how it is to be tested, when it is to be tested, and how test results fit into the general evaluation plan. Test items should be written according to established guidelines. They should be completely free from ambiguities, clearly produced, and clearly copied. They should clearly convey to students the type of response they are to give. Time should not be a discriminating factor, and thinking time should be allowed for when tests are constructed. When students experience failure in testing situations, some consideration should be given to providing an extra chance for success, such as an opportunity to write a parallel form, especially if the failure occurs early in the term.

Although test results may carry considerable weight when evaluation decisions are made, these decisions should not be made on the basis of test results only. Averaging of test results does not provide a valid picture of student achievement in subject areas where conceptual development and skill acquisition are sequential. Often, extra information will be needed to make a decision that is in the best interest of the student, since testing cannot be regarded as comprehensive evaluation.

Results of many tests, particularly those having selected-response items, provide information only on product. Testing is, therefore, but a subset in the domain of evaluation.

Guidelines

1. Use more than one data source in evaluating student progress.
2. Conduct observations regularly and systematically.
3. Where official records are required, the results of observations should be documented through the use of checklists, rating scales, anecdotal records and running records.
4. Allow primary and elementary students to complete projects at school where teachers can provide the necessary assistance and guidance.
5. When assigning projects, provide students with a written guide containing the intended outcomes of the project, a description of the task, relevant dates, references and evaluation criteria.
6. The value of items testing particular outcomes on a test should be in proportion to the importance of those outcomes.
7. Give proper attention to test validity and reliability. Use of publishers' tests should be restricted to tests that contain properly constructed items which measure the outcomes outlined in curriculum guides.
8. Ensure students know what is to be tested, how and when it is to be tested, and how test results fit into the general evaluation plan.
9. Tests should not be used as a control mechanism.
10. Teach students how to write tests.
11. Construct test items according to established guidelines.
12. Ensure tests are clearly produced, clearly copied and contain no unfamiliar item types. Clear up any real or perceived difficulties students might experience while writing the tests.
13. Select test items on the basis of how well they measure course outcomes, not on the basis of difficulty or how well they discriminate among students. Discriminate at the higher levels of achievement in such a way other students are not penalized.
14. When testing, use items requiring different cognitive processes.
15. When scheduling tests, allow for slightly more time than would be necessary for the average student to complete the test.

16. Require students to complete only those tasks which would be reasonable under the constraints of a time limit.
17. Allow for thinking time when constructing tests.
18. Use all evaluation information and exercise professional judgement to obtain a summative result rather than compute an average of all measurements gathered over the course of instruction.
19. Allow a student who passes a final, comprehensive test in a course that is amenable to measurement to pass that course.
20. Where mid-term tests are conducted, they should be administered in the classroom rather than during a block of time set aside for that purpose.
21. Block end-of-year testing should be restricted to senior high grades.

Grading

Policy Statement For summative evaluation, grades will indicate performance in relation to the stated outcomes.

Rationale

Summative evaluation is an attempt by a teacher to give an assessment of a student's achievement of outcomes after considering all relevant data gathered from a variety of sources. The challenge for the teacher is to integrate evaluation data from all sources to provide a grade in relation to course outcomes. When more than one type of outcome, such as cognitive and affective are being evaluated, a separate grade should be used for each.

When teachers provide a summative grade, it should be realized there is still a formative dimension. Teachers, as well as parents and students, are expected to use the information for future decisions about the teaching-learning process. The provision of descriptive comments can identify a student's strengths and pinpoint special needs that may require remedial attention. However, there is still a need to use some grading scheme, ranging from broad categories at the primary level to specific percentage grades for senior high courses.

Guidelines

1. Grades should always be referenced to the stated outcomes of a course or program.
2. The same grades given by different teachers using the same evaluation policy should have, at least approximately, the same meaning.
3. Even where percentage grades are used, an algorithmic approach to determining final grades should not be used.

Reporting

Policy Statement: *A system of formal and informal communication will be maintained in each school.*

Rationale

Parents have a right to know how their children are doing at school. If they are kept informed, they are in a position to have a positive influence upon their children's achievement. Communication can be qualitative or quantitative. Qualitative reports provide descriptive statements about achievement of outcomes. Quantitative reports of student achievement provide numerical grades. Communication can be one-way, two-way or multi-way. One-way communication occurs when schools send home reports, and the communication ends there. Two-way communication is most often established through teacher-parent/guardian conferences or where parents respond to report cards. Multi-way communication occurs where conferences are attended by teachers, parents and students and any others who have an interest.

Formal reporting is often infrequent in that it takes place only during designated times of the year. Informal networks should be set up so that information flows continuously between the home and the school. Much information can be exchanged through telephone calls, sending home annotated work samples, and so on. The best method to use is that which best serves the purpose of continuously providing the home with relevant information.

In most cases, students are exposed to more than one teacher, and no one is responsible for monitoring progress. Therefore, each student should be assigned to a monitoring teacher who would receive evaluation data from other teachers as soon as it becomes available. This teacher uses these data to provide timely feedback during informal contacts with the home.

Guidelines

1. Communicate the results of formative evaluation to students immediately if possible. Where students take a test, and immediate feedback is not possible, feedback should be provided within two or three instructional days.
2. Provide parents or guardians with a report of their children's achievement at least four times a year.
3. Provide for teacher-parent/guardian conferences immediately following the release of report cards.

4. Between formal reporting periods, communicate with the home informally.
5. In high schools, each student should be assigned to a monitoring teacher.
6. Report cards should contain quantitative and qualitative information.
7. For Kindergarten, there should be a planning meeting and a parent-teacher conference prior to the first report card being released. The report should consist of a checklist and anecdotal comments relating to strengths, needs and recommendations.
8. The primary report should consist of a checklist and anecdotal comments relating to strengths, needs and recommendations.
9. At the elementary level, a rating scale and anecdotal comments should be used on report cards.
10. At the intermediate level, numerical or letter grades accompanied by rating scales and anecdotal comments should be used to report achievement in each strand of a subject area. Overall achievement in each subject area should be indicated by a numerical grade. Grading should be referenced to the outcomes of the program.
11. At the senior high level, percentage grades and anecdotal comments should be used for progress reports.

Promotion

Policy Statement: Students will be promoted unless appropriate arrangements cannot be made in the next grade there is general agreement promotion is not in the students' best interest.

Rationale

Retention in a grade is beneficial for only a very small percentage of students. The identification of these students is a difficult task. A decision to retain should be made only when no satisfactory accommodation can be made for the student in the next grade. Such a decision should be made by a panel of people rather than an individual. As a rule of thumb, no student should be retained for failing to achieve satisfactorily the outcomes of a course in which learning is not sequential and cumulative.

Guidelines

1. Retain students only when satisfactory accommodation cannot be made in the next grade. When a student is retained, a modified program should be put in place.
2. Before a decision is made to retain a student in the same grade, there should be a consultation process involving the school administration, teachers, the parents or guardians of the student and any other persons who have an interest.
3. Retain students only on the basis of achievement in courses in which learning is sequential and cumulative.

APPENDIX A: LEVELS OF THE COGNITIVE DOMAIN WITH APPROPRIATE VERBS

(Adapted from Gronlund , 1985)

Knowledge: The ability to recall learned materials. It can range from the recall of simple facts to complete theories. It represents the lowest level of learning outcome, requiring only that the student recall previously learned information. Statements indicating what students should be able to do could use such terms as:

knows	lists	matches	what
defines	names	outlines	where
identifies	reproduces	states	when
describes	selects	recognizes	who
labels	recalls	memorizes	

An example of a knowledge item is shown below.

Outcome: Students will match chemical elements with their symbols.

Match each chemical element found in column I with its correct symbol found in column II. Place the number which corresponds with the correct answer in the blank provided to the right of column

I		II
1. Copper		1. Ag ___
2. Gold		2. Au ___
3. Lead	3. Cu ___	
4. Mercury		4. Hg ___
5. Sodium		5. K ___
6. Silver		6. Na ___
7. Tin		7. Pb ___
		8. Si ___
		9. Sn ___

Comprehension: The ability to grasp the meaning of material learned. The student may show understanding of the material by translating it from one form to another, by conveying meaning, or by making summary statements about it. Action terms used as indicators of student achievement include the following:

understands	interprets	translates	shows
estimates	justifies	converts	distinguishes
defends	explains	generalizes	extends
infers	predicts	paraphrases	gives examples
rewrites	summarizes	demonstrates	why
how	condenses		

An example of a Comprehension item is shown below.

Outcome: Students will interpret information plotted on a line graph.

Which month has the lowest average precipitation? (Graph not provided)

- (A) January
- (B) February
- (C) May
- (D) December

Application: Students are required to use learned materials in new and concrete situations. Indicators of student achievement would include the following:

applies	organizes	solves	relates
constructs	changes	computes	produces
demonstrates	discovers	manipulates	uses
modifies	operates	predicts	

An application item appears below.

Outcome: Students will apply the principle of supply and demand to particular situations.

“If the demand for a commodity or service is relatively constant, a decrease in its supply will result in an increase in its market value.” Read each of the following three statements, and place a check mark next to any which agrees with this principle.

1. Since World War II, there has been an upward trend in the price of stocks.
2. Fresh fruit and vegetables cost more when not in season.
3. The cost of medicine is higher now than it was ten years ago.

Analysis: The ability to break material down into its component parts so that its organizational structure may be understood. The student demonstrates attainment of outcomes through the ability to identify parts, show relationships, and recognize organizational principles. Outcomes written at this level would use words such as the following:

recognizes	categorizes	differentiates	relates
diagrams	distinguishes between	classifies	outlines
illustrates	discriminates	analyzes	

An analysis item appears below.

Outcome: Students will analyze a poem to determine its tone.

What is the tone of the poem? (poem not provided)

- (A) jaunty
- (B) sarcastic
- (C) silly
- (D) sorrowful

Synthesis: The ability to put parts together to form a new whole. The student demonstrates an ability to devise a new plan of operation, or to produce a set of abstract relations. Synthesis outcomes cannot be tested with items that require students to select a response. Students must actually create their own responses where the ability to synthesize is being evaluated. The emphasis is on creativity, and appropriate verbs would include the following:

proposes	integrates	creates
categorizes	combines	compiles
devises	rearranges	reconstructs
composes	formulates	develops

An example of a synthesis items appears below.

Outcome: Students will show how various dimensions of a play are interrelated.

Identify two characters from any Shakespearean play, and in a multi-paragraph essay, explain how these characters contribute to the development of a theme in the play.

Evaluation: The ability to judge the value of material. The student might be required to judge the value of a statement, a piece of prose, a poem, an advertisement or a research report. Evaluation outcomes cannot be tested with items that require students to select a response. Students must actually create their own responses where the ability to evaluate is being assessed. To indicate that the outcome is achieved, the student may be required to:

judge	appraise	conclude
contrast	criticize	justify
interpret	support	evaluate

An example of an evaluation item appears below.

Outcome: Students will evaluate the emergence of modern phenomena as the result of the forces of cultural geography.

“Singapore’s location alone fully explains its growing importance as a transshipment hub.”
With reference to the factors that give rise to transshipment centres, assess the accuracy of this statement.

The use of appropriate verbs can be further illustrated as follows:

TOPIC: GRAPHS

- Outcome: Students will demonstrate an understanding of graphs.
- Knowledge: Students can identify the information to be graphed.
- Comprehension: Students can interpret information given in a graph.
- Application: Students can use information to construct a graph.
- Analysis: Students can analyze information given in a graph.
- Synthesis: Students can make summary statements or draw conclusions based on information given in the graph.
- Evaluation: Students can judge the adequacy with which conclusions are supported by available data.

APPENDIX B: LEVELS OF THE AFFECTIVE DOMAIN WITH APPROPRIATE VERBS
 (Adapted from Gronlund, 1985)

Affective Outcome: The student will demonstrate an interest in Science.

- Specific Indicators:**
1. The student participates in extra-curricular activities such as science clubs and science fairs.
 2. The student does extra readings in areas related to science.
 3. The student watches and discusses television programs related to science.

Five levels: examples and appropriate verbs.

Receiving: A willingness to attend to specific stimuli.

- | | | |
|-----------|-----------|-----------|
| to accept | to listen | to choose |
| to select | to ask | to attend |

Responding: A willingness to participate.

- | | | |
|------------|-----------|-----------|
| to answer | to select | to write |
| to conform | to tell | to follow |

Valuing: The worth or value a student attaches to a specific object, behaviour, or phenomenon.

- | | | |
|------------|-----------|---------------|
| to choose | to invite | to share |
| to support | to join | to appreciate |

Organization: Concerned with bringing together different values, resolving conflicts between them, and beginning the construction of an internally consistent value system.

- | | | |
|--------------|-----------|-----------------|
| to formulate | to relate | to defend |
| to abstract | to define | to put in order |

Characterization: Deals with patterns of adjustment and is characterized by pervasiveness, consistency, and predictable behaviour.

- | | | |
|-----------------|-------------|-----------|
| to discriminate | to complete | to behave |
| to practise | to serve | to verify |

APPENDIX C: LEVELS OF THE PSYCHOMOTOR DOMAIN WITH APPROPRIATE VERBS

(Adapted from Gronlund, 1985)

Perception: The use of sense organs to obtain cues that guide motor activity.

Outcome: The student can relate music to a particular dance step.

Verbs: chooses describes detects
separates identifies differentiates

Set: Readiness to take a particular type of action.

Outcome: The student demonstrates proper bodily stance for batting a ball.

Verbs: begins displays explains
moves proceeds reacts
responds shows starts

Guided Response: Deals with the early stage in learning a complex skill. It includes imitation as well as trial and error.

Outcome: The student can perform a golf swing as demonstrated.

Verbs: assembles builds constructs
dismantles dissects fastens
fixes measures manipulates
mends sketches

Mechanism: Deals with performance acts where the learned responses have become habitual and the movements can be performed with some confidence and proficiency.

Outcome: The student writes smoothly and legibly.

Verbs: The same list as for guided response.

Complex Overt Response: Skillfully performs a motor act that requires complex movement patterns.

Outcome: The student can demonstrate correct form in swimming backstroke.

Verbs: The same list as for guided response.

/

Adaption: Deals with skills which are so well developed the individual can modify movement patterns to fit special requirements or to meet a problem situation.

Outcome: The student modifies his swimming stroke to fit the roughness of the water.

Verbs: adapts alters changes
 varies rearranges reorganizes
 revises

Origination: Deals with creating new movement patterns to fit a particular situation or specific pattern.

Outcome: The student creates a new dance step.

Verbs: arranges combines composes
 constructs creates designs
 originates

APPENDIX D: CHECKLIST

	Rarely	Sometimes	Usually	Comments
<p><i>Reading Strategies</i></p> <p>Expects the text to make sense Searches for meaning Makes predictions Self-corrects Processes chunks of language</p> <p><i>When in Difficulty:</i></p> <p>Reads on to end of sentence Re-reads sentence Guesses using initial letter as cue</p> <p><i>After Reading:</i></p> <p>Can re-tell story in own words Able to discuss characters Recalls main ideas Understands developing sequence Can describe setting</p>				

APPENDIX E: ANECDOTAL RECORD

Name:

Date: _____

Grade: _____

Observed Behaviour	Comments	Program Plans

APPENDIX F: RATING SCALE

MUSIC

Evaluation of Recorder Playing

Name: _____ Grade: _____ Date: _____

RHYTHM

Correct rhythm:

Consistent tempo

FINGERING AND NOTES

Correct fingering and notation

TONGUEING AND BREATHING

Correct phrasing and articulation

POSTURE

Stance and holding of recorder

OVERALL PERFORMANCE

Key: 2 - good (but not necessarily perfect)

1 - satisfactory

0 - unsatisfactory

APPENDIX G: RUNNING RECORD (Source: Wortham, S.C.)

<p>Child's Name: <u>Christopher</u> Age: <u>4</u> Location: <u>KinderKare</u> Date and Time: <u>June 21, 1988</u> <u>8:40-9:10</u> Observer: <u>Cindy</u> Type of Development Observed: <u>Social and Cognitive</u></p>	
Observation	Notes or Comments
<p>Chris is playing with a toy. He says, "Kelly, can I keep it?" several times until he gets an answer. He moves on to a toy guitar and plays it while he supervises the other children by walking around the room. He tells everyone to sit down at the tables after the teacher says to.</p> <p>Chris sits by a friend and talks about eating granola bars. He watches and listens to the conversation on either side of him. He's still unaffected by the loud temper tantrum of another child. Then he Notices her and watches. He tries to explain this behaviour to the others by saying a plant was split.</p> <p>He follows teacher's directions. Then he decides he wants to be in on a secret. A boy shoves him away. Chris informs him that he <i>can</i> hear if he wants to. This has caused him to disobey the teacher. He has to sit out of the circle. He walks over to the chair, sits down, gets up immediately, comes back to the circle undetected by the teacher. He joins the circle.</p>	<p>Chris is polite to others.</p> <p>Chris is helping his classmates to follow the rules.</p> <p>Chris is interested in what others have to say.</p> <p>Chris tries to make sense of a child's behaviour.</p> <p>Chris chooses appropriate ways assert himself.</p>

APPENDIX H: PROJECT ASSIGNMENT SHEET

Topic Choice(s):

1. Assign a specific topic.
2. Provide a class of topics from which students can make a selection.
3. Allow students a propose their own topics from within a general subject area.

Outcomes:

Provide a limited list of primary/specific outcomes concerning the learning expectations for students. These outcombjectives should be stated in terminology which is familiar to students.

Task Description:

Include in the task description:

1. what students may/may not do;
2. how they might do it;
3. where work will be completed, and
4. what forms are acceptable for a completed project.

Relevant Dates:

1. Preliminary dates:
 1. provide several dates for teacher/student discussion;
 2. provide dates for peer discussion and consultation, and
 3. provide dates for progress reporting.
2. Final submission date

Evaluation Criteria:

Include the specific criteria on which the project will be evaluated. Indicate the weights assigned to each criterion.

APPENDIX I: STUDENT SELF-EVALUATION FORM

Am I developing into an independent reader?

Name: _____

Grade: _____

Date: _____

Choosing Books

- Do I choose at the correct level?
 - Do I choose a variety of books?
 - Do I listen to the suggestions of others?
 - Do I use all available sources?
-

Reading Independently

- Do I enjoy silent reading time?
 - Do I choose to read at other times?
 - Do I read different books for different purposes?
 - Do I know what to do when I don't understand something?
 - Do I know what to do when I don't know a word?
-

APPENDIX J: TABLE OF SPECIFICATIONS

	Conceptual	Procedural	Problem-Solving
Numbers & Numerations	26	28	
Geometry	10	0	20
Measurement	6	4	
Statistics	6	0	
Total	48	32	20

APPENDIX K: TEST BLUEPRINT

	CONCEPTUAL	PROCEDURAL	PROBLEM-SOLVING
NUMBER			10
Whole	1		
Fractional	4		
Decimal	3		
Number Theory	2		
Percent	2	1	
Op/Prop (M,S)		2	
Op/Prop (M,D)	1	4	
Op Dec (A,S)		2	
Op Dec (M,D)		5	
GEOMETRY			
Solid	1		
Plane	4		
MEASUREMENT			
Linear	1		
Area		1	
Volume		1	
Cap/mass/time	2		
STATISTICS			
Tab/char/gra	3		

APPENDIX L: CONSTRUCTED-RESPONSE ITEMS

Extended-response items: The examinee responds to a question with a paragraph or more. These items are often referred to as essay items.

Guidelines for Construction:

1. Base questions on the outcomes of the unit of work being tested.
2. Questions requiring higher-order mental processes can not be a repetition of situations or problems used during the instruction.
3. Wording of the question should be specific and clear. Examinees should know exactly what they are to do.
4. Note on the examination the number of marks that each essay item is worth. Marks allotted should reflect the difficulty of the item and the time required answer it.
5. To obtain adequate reliability, it is necessary to prepare a scoring guide beforehand.
6. Examinees should be made aware beforehand of the criteria against which their essays will be evaluated.

Restricted-response items: The examinee responds to a question with a word, sentence, a brief paragraph or a list of statements. This includes the shorter essay items or completion objectives.

Guidelines for Construction: In addition to those stated above which may apply to many restricted-response items, the following should be noted for completion items.

1. Items should require a single-word answer rather than a phrase. Synonyms should be keyed as correct.
2. Only key words should be left blank.
3. The blank should appear at, or very near, the end of the statement. This ensures that the students will have all necessary information required to provide the answer. It also facilitates scoring.
4. No cues to the answer should be provided. If the word an appears before a blank, for example, it cues the student that the answer begins with a vowel. Therefore a/an should be used.

5. Where units are required, they should be put in.
6. Statements should not be lifted directly from the text, since doing this encourages rote memorization.
7. One element only should be omitted from a statement.
8. All blanks should be of uniform length. A single blank should always be used even when the answer is two words such as New York.
9. Statements should be definite such that they cannot have more than one correct answer.

APPENDIX M-1: SELECTED-RESPONSE ITEMS

True-False

Description: The examinee indicates whether a statement is true or false.

General: These items are best suited for testing knowledge and are very difficult to write for some subjects such as social sciences. Tests comprised of true-false items do not produce reliable results, but reliability can be improved by including many items. Another limitation is that it has no diagnostic function. They are best when only two choices exist such as right or left, distinguishing between cause and effect and distinguishing fact from opinion.

Guidelines for construction:

1. Avoid absolutes like *all* and *never*. Statements containing such words are usually false, and the use of qualifiers cues the examinee to answer true.
2. Key approximately one-half the items true.
3. Avoid making statements false by insertion of the word *not*. If a negative has to be used, the negative should be highlighted.
4. Avoid double negatives.
5. Make sure that there are no cues for students (such as true statements always being longer than false ones).
6. Statements should be short, and embody one idea.
7. Avoid taking statements directly from a text.
8. If opinion is used, it should be attributed to the source.
9. Statements must be clearly true or clearly false.
10. Statements must not include indefinite terms, degrees, or amounts.
11. To make false statements sound true, phrases which give them a *ring of truth* may be used.

APPENDIX M-2: SELECTED-RESPONSE ITEMS

Matching

Description: This format contains two parallel columns with each word, phrase, number or symbol in one column being matched with a word or phrase in the other column.

General: Matching items are limited to measuring knowledge which is based on simple associations. They are appropriate where an outcome requires “the ability to identify the relationships between two things”. Thus, they are restricted to relatively small areas of achievement. Matching items are thought to be easy to write, but they are often made difficult by the requirement that each response be a plausible answer for each premise. The advantage of matching items is that they can measure knowledge of much content in a short time.

Guidelines for construction:

1. The lists should have more responses than premises.
2. Responses should be placed in some logical order (alphabetical, for example).
3. Each alternative must be a plausible solution to each premise.
4. Instructions should be clearly stated. If responses may be used more than once, it should be indicated.
5. All items should be on one page.
6. Responses should be as concise as possible. Verbatim statements from the text should not be used.
7. Difficulty level may be controlled by the discriminations the examinee is expected to make. Generally speaking, the broader the terms in the premises list, the easier the item is.
8. The responses and premises should contain no cues such as plural forms.
9. There should be 8-10 matching items in one set.
10. Both columns should be labelled.

APPENDIX M-3: SELECTED-RESPONSE ITEMS

Multiple-choice

Description: The student indicates which one of three or more choices is correct or most correct.

General: Multiple-choice items can be used to measure more than recall of knowledge. (See earlier discussion in this section). These items are adaptable to most subjects. If they are well-written, they are easily understood by students, and they are easy to score.

Guidelines for construction:

1. The stem should be a complete statement about the problem to be solved.
2. The stem should contain no unnecessary or irrelevant information.
3. The stem should be stated in clear and precise language.
4. The stem should contain as much of the item as possible such that information will not have to be repeated in each alternative.
5. Negative statements should be avoided if possible. If they cannot be avoided, the negative should be placed near the end of the stem and highlighted.
6. All distractors should be consistent with the grammatical and syntactical construction of the stem.
7. Distractors should be approximately equal in length.
8. If an opposite is used as one distractor, the other two distractors should be opposites as well.
9. All distractors should be plausible answers to the problem presented in the stem.
10. There should be no grammatical cues to the right answer.
11. One alternative should be clearly right and the distractors clearly wrong except in the case of the best answer type.
12. Avoid using camouflage distractors (none of the above, all of the above).

13. Modifiers such as sometimes and usually should be avoided.
14. If the correct answer has important-sounding words, distractors should contain such words as well.
15. Choices should be listed vertically and on separate lines.
16. Letters should be in front of choices.
17. Statements should not be lifted directly from the text.
18. Items should contain no tricks
19. Items that test higher-order mental processes must have an element of novelty, but situations presented in the item should be similar to those used in class.
20. Use the *best* answer-type only when there is no correct answer.

REFERENCES CITED

- American Federation of Teachers, National Council on Measurement in Education & National Education Association. (1990). Standards for teacher competence in educational assessment of students.
- Anderson, J. D. & Bachor, D. G. (1993). Assessment practices in the elementary classroom: Perspectives of stakeholders. (ERIC Document Reproduction Service No. ED 361 394)
- Anderson, L., Evertson, C., & Brophy, J. (1982). Principles of small-group instruction in elementary reading (Occasional Paper No. 58). East Lansing: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Service No. ED 223 981)
- Anderson, S. A. and the Amerman PTA, Amerman Elementary School (1993). School-home communications: a qualitative survey. (ERIC Document Reproduction Service No. ED 364 349)
- Baker, E. L. (1998). Model-Based performance assessment. Technical Report 465. Los Angeles, CA: Centre for the Study of Evaluation, University of California, Los Angeles
- Baker, E. L., O'Neil, H. F. & Linn, R. L. (1993). Policy and validity prospects for performance-based Assessment. American Psychologist, Vol 48, No.12.
- Bloom, B. S. (1956). Taxonomy of educational objectives: Handbook I, the cognitive domain. New York: Longmans, Green.
- Brophy, J. (1981). Teacher praise: A functional analysis. Review of Educational Research, 51(1), 5-32.
- Brophy, J., & Evertson, C. (1976). Learning from teaching: A developmental perspective. Boston: Allyn & Bacon.
- Brophy, J., & Good, T. (1986). Teacher behaviour and student achievement. In M. Wittrock (Ed.), Handbook of research on teaching (pp.328-375). New York: Macmillan.
- Bursuck, W., Polloway, E. A., Plante, L., Epstein, M. H., Madhavi, J., & McConegy, J. (1996). Report card grading and adaptations: A national survey of classroom practices. Exceptional Children. Vol 62, No. 4. pp 301-308.
- Canadian Education Association. (1989, March). Grade promotion and retention: practices in Canadian School boards. CEA Information Note, Toronto, ON.
- Chansky, N. M. (1975) A critical examination of school report cards from K through 12. Reading Improvement, 12, (3).

Denham, C., & Lieberman, A. (Eds.). (1980). Time to learn. Washington, DC: National Institute of Education.

Dennebaum, J. M. & Kulberg, J. M. (1994) Kindergarten Retention and Transition Classroom: Their Relationship to Achievement. Psychology in the Schools, 31, 1, 5-12.

Dietel, R. (1993). What works in performance assessment? Proceedings of the CRESST conference. Sept. 10-12. (Document Reproduction Service No.ED 362 562.)

Dobbs, V. & Neville, D. (1967). The Effects of Nonpromotion on the Achievement of Groups Matched from Retained First Graders and Promoted Second Graders. Journal of Educational Research, 60, 472-475

Ebel, R. (1972). Essentials of educational measurement. Englewood Cliffs, NJ : Prentice-Hall.

Eisner, E. (1985). The art of educational evaluation. Philadelphia: Falma, Taylor & Francis.

Eiss, A.. & Harbeck, M. (1969). Behavioral objectives in the affective domain. Washington, DC: National Science Supervisors Association.

Emmer, E., Evertson C., & Brophy, J. (1979). Stability of teacher effects in junior high classrooms. American Educational Research Journal, 16(1), 71-75.

Friedman, S. J. and Frisbie, D. A. (1995) The influence of report cards on the validity of grades reported to parents. Educational and Psychological Measurement, 55, (1).

Frisbie, D. A. and Waltman, K. K. (1992) Developing a personal grading plan. Educational Measurement: Issues and Practices, 6(4) 29-37

Frost, J. (1989). Assessing program and instruction in early primary grades. The Canadian School Executive, 8 (7), 11-15.

Gagné, R. M. The conditions of learning and theory of instruction. (4th ed.). New York: Holt, Rhinehart & Winston.

Good, T., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. Journal of Teacher Education, 28(3), 49-53.

Gronlund, N. E. (1985). Measurement and evaluation in teaching (5th ed.). New York: Macmillan.

Gronlund, N. E. (1995). How to write and use instructional objectives (5th ed.). Englewood Cliffs: Prentice Hall Inc.

Guillermo, S. & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical Issues. Educational Measurement: Issues and Practice Fall, 1997.

Guskey, T. R. (1996). Introduction. ASCD Yearbook, 1996: Communicating Student Learning. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Hall, K. (1990) Determining the success of narrative report cards. Unpublished Report. Curry School of Education. University of Virginia.

Herman J. (1997) Assessing new assessments: How do they measure up? Theory Into Practice Vol. 36, No. 4.

Herman, J. L., Klein, D. C. D., Heath, T. M. & Wakai, S. T. (1994). A first look: Are claims for alternative assessment holding up? (Technical Report 39)1. Los Angeles, CA: Centre for the Study of Evaluation, University of California, Los Angeles.

Hiebert, E. H. (1991). Teacher-based assessment of literacy learning. In J. Flood, J. M. Jensen, D. Lapp, & J. R. Squire (Eds.) Handbook of research on teaching the English language arts (pp.510-520). Toronto: Macmillan Publishing Company.

Hoffman, P. J. (1995). The challenge of revising report cards. In Costa, A. L. & Kallick, B. (Eds.) Assessment in the Learning Organization: Shifting the Paradigm (ERIC Document Reproduction Service No. ED 395 932)

Hughes, D. (1973). An experimental investigation of the effects of pupil responding and teacher reacting on pupil achievement. American Educational Research Journal, 10(1), 21-37.

Krathwohl, D., Bloom, B. S., & Masia, B. (1964). Taxonomy of educational objectives: Handbook II, the affective domain. New York: David McKay.

Kundert, D. K., May, D. C & Brebt, D. (1995) A Comparison of Students Who Delay Kindergarten Entry and Those Who are Retained in Grades K-5. Psychology in the Schools, 32, 3, 202-209.

Lake, K. and Kafka, K. (1996) Reporting methods in grades k-8 ASCD Yearbook, 1996: Communicating Student Learning. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Liaw, F. & Meisels, S. J. (1993). Failure in grade: Do retained students catch up? Journal of Educational Research, 87, 2, 69-77.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. Paper presented at the annual meeting of the American Psychological Association. Washington, D. C., August

Linn, R. L., Baker, E. L. and Dunbar, S. B. (1991) Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher. November.

Lotz, M. & Sukorsky, J. (1989). Parents' and teachers' attitudes towards progress reporting conferences. (ERIC Document Reproduction Service No. ED 306 264)

Lovelady, I. H. (1991) A model parent involvement initiative. A paper presented at the 21st annual conference of the National Black Child Development Institute. St. Louis, Missouri, October 23-25. (ERIC Document Reproduction Service No. ED 353 034)

Messick, S. (1989). Validity. In R. L. Linn (Ed.) Educational Measurement (3rd ed., pp. 13-103). Washington, DC: American Council on Education & National Council on Measurement in Education.

Newfoundland Department of Education. (1998). Pathways to Programming and Graduation.

Newfoundland Department of Education. (1998). Programming for individual needs: Assessment and evaluation.

Nitko, A. J. (1996). Educational assessment of students. Englewood Cliffs: Prentice-Hall, Inc.

Norris, S. P. (1992). Assessing thinking in the classroom: Building on the curriculum. St. John's: Memorial University of Newfoundland.

Novak, J. R. Herman, J. L. & Gerhart, M. (1996). Issues in portfolio assessment: The scorability of narrative collections. (Technical Report 410). Los Angeles, CA: Centre for the Study of Evaluation, University of California, Los Angeles

Nuttall, D. L. (1992). Performance assessment: The message from England. Educational Leadership. Vol 49.

Olhausen, M. M., Powell, R. R. and Reitz, B. S. (1994) Parents' views of traditional and alternative report cards. The School Community Journal. Vol. 4, No. 1. Spring/Summer, 1994.

Oosterhof, A. (1994). Classroom applications of educational measurement. Toronto: Maxwell Macmillan Canada.

Paris, S. G. & Ayres, L. R. (1994). Becoming reflective students and teachers with portfolios and authentic assessment. Washington: American Psychological Association.

Pawlas, G. E. (1985) Improving public support: The answer is communications. Childhood Education. 61(4), March/April.

Perkins, J. & Buchanan, A. (1983). How parents find out about student progress. (ERIC Document Reproduction Service No. ED 250 787)

Pomplun, M. (1988) Retention: The Earlier, the Better? Journal of Educational Research, 81, 5, 281-87.

Popham, W. J. (1999). Classroom assessment: What teachers need to know. Needham Heights, MA: Allyn & Bacon.

Quellmalz, E. S. (1985). Developing reasoning skills. In J. R. Baron & R. J. Sternberg (Eds.), Teaching Thinking Skills: Theory and Practice (pp. 86-105). New York: Freeman.

Resnick, L. B. & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), Changing Assessments: Alternative Views of Aptitude, Achievement and instruction. Boston: Kluwer Academic Publishers.

Rogers, W. T. (1996). Code of professional responsibilities in educational measurement. Assessment in Education: Principles, Policy & Practice Vol.3, No. 3 (pp 401-411).

Roid, G., & Haladyna, T. (1982). A technology for test item writing. New York: Academic Press.

Scott, L. (1992). Improving evaluation of third grade literacy using authentic techniques and self-assessment. (ERIC Document Reproduction Service No. ED 347 510)

Shaeffer, M. & Hook, J. (1993) Are Extra-Year Classes Worth It? American School Board Journal, 180, 31-32.

Shepard, L. A. (1997) Insights Gained from a Classroom-Based Assessment Project. Technical Report 451. Los Angeles, CA: Centre for the Study of Evaluation, University of California, Los Angeles.

Shepard, L. A. & Bliem, C. L. (1995). An analysis of parent opinions and changes in opinions regarding standardized tests, teacher's information and performance assessments. (Technical Report 397). Los Angeles, CA: Centre for the Study of Evaluation, University of California, Los Angeles.

Simpson, E. (1972). The classification of educational objectives in the psychomotor domain: Vol. 3, the psychomotor domain. Washington, DC: Gryphon House.

Stephens, L. S. (1980). Using parental preferences in report cards to design a report card for Blennerhassett kindergarten. (ERIC Document Reproduction Service No. ED 195 350)

Stiggins, R. J. (1997). Student-Centered Classroom Assessment. (2nd ed.) New Jersey: Prentice-Hall, Inc.

Tanner, K. T. (1993) Student Retention Policy: The Gap between Research and Practice. Journal of Research in Childhood Education, 8, 1, 69-77.

Tikunoff, W., Berliner, D., & Rist, R. (1975). An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample (Technical Report No. 75-10-5). San Francisco: Far West Laboratory. (ERIC Document Reproduction Service No. ED 150 110)

Waltman, K. K. and Frisbie, D. A. (1994) Parents understanding of their children's report card grades. Applied Measurement in Education, 7, (3).

Wiggins, G. (1996) Honesty and fairness: Toward better grading and reporting ASCD Yearbook, 1996: Communicating Student Learning. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Wise, L. L. (1992) Lessons learned from military performance assessment. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco CA, Apr 21-23. Eric Document ED 346 165.

Wortham, S. C. (1990) Tests and Measurement in Early Childhood Education. Columbus, Ohio: Merrill Publishing Company.