TP 13589E

ASSESSMENT OF FITNESS-FOR-DUTY TECHNOLOGIES IN TRANSPORT OPERATIONS

Prepared for

Transportation Development Centre
Transport Canada

by

Rhodes & Associates Inc.

April 2000

TP 13589E

Assessment of Fitness-for-Duty Technologies in Transport Operations

by

Wayne Rhodes
Rhodes & Associates Inc.
and
Alex Vincent
Transportation Development Centre
Transport Canada

April 2000

Un sommaire français se trouve avant la table des matières.

Transport Canada / Transports Canada

PUBLICATION DATA FORM

1. Transport Canada Publication No. TP 13589E	2. Project No. 9758	3. Recipient's Catalogue No.
4. Title and Subtitle Assessment of Fitness-for-Duty Technologies in Transport Operations		5. Publication Date April 2000
		6. Performing Organization Document No.
7. Author(s) W. Rhodes (Rhodes & Associates Inc.) and A. Vincent (Transportation Development Centre)		8. Transport Canada File No. ZCD2450-209
9. Performing Organization Name and Address Rhodes & Associates Inc. 177 Jenny Wrenway Toronto, Ontario M2H 2Z3		10. PWGSC File No. XSD-9-01439
		11. PWGSC or Transport Canada Contract No. T8200-9-9549
12. Sponsoring Agency Name and Address Transportation Development Centre (TDC) 800 René Lévesque Blvd. West Suite 600 Montreal, Quebec H3B 1X9		13. Type of Publication and Period Covered Final
		14. Project Officer Alex Vincent
15. Supplementary Notes (Funding programs, titles of related publications, etc.)		

16. Abstract

A wide range of fitness-for-duty (FFD) and readiness-to-perform (RTP) testing systems (technologies) were reviewed and assessed for potential application in transport operations. Major research papers and descriptions of FFD/RTP systems appropriate for application to transport operations were identified, using a rating system based on the following factors:

- past application to the transportation industry;
- validation with third-party validity data;
- test session duration;
- ability to produce quick results (immediate versus longer duration due to analytical requirements);
- potential reliability (i.e. a self-contained, field-proven system versus experimental lab-based equipment); and
- appropriateness of test elements to transport operations.

The reviewed systems are listed and described according to these factors. Recommendations include a table of optimal FFD/RTP testing systems.

17. Key Words Fitness-for-duty (FFD), readiness-to-perform (RTP), operator testing systems	18. Distribution Statement Limited number of copies available from the Transportation Development Centre E-mail: *tdccdt@tc.gc.ca*

19. Security Classification (of this publication)	20. Security Classification (of this page)	21. Declassification (date)	22. No. of Pages	23. Price
Unclassified	Unclassified	—	xii, 48	Shipping/ Handling

Transports Canada | Transport Canada

FORMULE DE DONNÉES POUR PUBLICATION

1. N° de la publication de Transports Canada TP 13589E	2. N° de l'étude 9758	3. N° de catalogue du destinataire
4. Titre et sous-titre Assessment of Fitness-for-Duty Technologies in Transport Operations		5. Date de la publication Avril 2000 6. N° de document de l'organisme exécutant
7. Auteur(s) W. Rhodes (Rhodes & Associates Inc.) et A. Vincent (Centre de développement des transports)		8. N° de dossier - Transports Canada ZCD2450-209
9. Nom et adresse de l'organisme exécutant Rhodes & Associates Inc. 177 Jenny Wrenway Toronto, Ontario M2H 2Z3		10. N° de dossier - TPSGC XSD-9-01439 11. N° de contrat - TPSGC ou Transports Canada T8200-9-9549
12. Nom et adresse de l'organisme parrain Centre de développement des transports (CDT) 800, boul. René-Lévesque Ouest Bureau 600 Montréal (Québec) H3B 1X9		13. Genre de publication et période visée Final 14. Agent de projet Alex Vincent

15. Remarques additionnelles (programmes de financement, titres de publications connexes, etc.)

16. Résumé

Un large éventail de systèmes (et appareillages) utilisés pour apprécier l'aptitude au travail (FFD, *Fitness-for-Duty*) et l'aptitude à la tâche (RTP, *Readiness-to-Perform*) ont été passés en revue, afin de déterminer leur applicabilité au secteur des transports. Les chercheurs ont dépouillé les dossiers descriptifs des systèmes ainsi que des rapports techniques, et ont coté ceux-ci en fonction des critères suivants :

- application antérieure dans le secteur des transports;
- validation à l'aide de données indépendantes;
- durée de la séance de test;
- délai d'obtention des résultats (immédiatement à l'issue du test ou au terme d'une période de traitement);
- fiabilité potentielle (système autonome, éprouvé en service, par rapport à un système expérimental, utilisé en laboratoire);
- adéquation entre les éléments du test et les activités de transport.

On trouve dans le rapport la liste des systèmes recensés et leurs caractéristiques, eu égard aux critères ci-dessus. La section des recommandations comporte un tableau des systèmes jugés optimaux.

17. Mots clés Aptitude au travail, aptitude à la tâche, systèmes de tests pour conducteurs	18. Diffusion Le Centre de développement des transports dispose d'un nombre limité d'exemplaires. Courriel : *tdccdt@tc.gc.ca*

19. Classification de sécurité (de cette publication) Non classifiée	20. Classification de sécurité (de cette page) Non classifiée	21. Déclassification (date) —	22. Nombre de pages xii, 48	23. Prix Port et manutention

CDT/TDC 79-005
Rev. 96

Canada

Executive Summary

This review was carried out to identify fitness-for-duty/readiness-to-perform (FFD/RTP) testing systems that have potential application to transportation operations. The review examined 26 systems. The sources for most descriptions were Gilliland & Schlegel (1993) and Miller (1996) and reports were available for many. Some systems were described only in company reports and were not as readily available. However, the most promising systems were well described and all data required for assessment were available.

The main factors considered for assessment were:

- Test components – what performance factors the test evaluates;
- Time required to complete the entire test;
- Resources required to conduct the testing;
- Past application to a transportation environment;
- Turnaround – when test results are available;
- Validity studies done on test components and measures; and
- Feasibility of applying the device to a typical transportation environment.

Selection of potential candidate FFD/RTP test systems and recommendations for implementation took the following needs into account:

- To define a specific FFD/RTP test within the application context;
- To establish a working theory on which the FFD/RTP testing must be based;
- To determine FFD/RTP goals to establish the criterion for measurement and its use;
- To carry out cross validation studies to ensure that the FFD/RTP tests show a true relationship to the chosen criterion;
- To undertake further research to assess the RTP test methods and to explore the fundamental principles on which the testing is based;
- To determine whether high face validity is desirable, particularly if that validity is not essential to a test criterion that predicts risk factors;
- To educate employees and management so that the need for face validity may be eliminated;
- To base tests on criterion validity for risk factors or work performance, or both;
- To use research results and discussions between vendors, employers, and employees for the basis of decisions about whether to focus on risk factor criterion validity and/or on work performance criterion validity;
- To ensure that test systems produce reliable results and have a high degree of availability;
- To use care and realistic expectations to determine the cut-off values for RTP tests;
- To handle failure to pass a test properly with adequate follow up and support;
- To realize that FFD/RTP testing is not a substitute or replacement for drug testing;
- To ensure adequate practice before an employee’s baseline can be established;
- To ensure that test duration is within practical limits given the operational environment;
- To assume that multiple tests may be needed for proper coverage;
- To ensure that tests contain elements relevant to the job even if they do not show face validity; and

- To ensure that the testing system is affordable to purchase, operate, and maintain.

Recommendations

The following research and implementation approaches are recommended:

- Profile the transportation jobs where safety is a major concern so that decisions can be made on the application of specific FFD/RTP testing systems;
- Determine the adequacy of appropriate testing systems for each job class and the level of customization of test components that may be required;
- Pilot test systems tailored to specific classes of jobs, such as psychomotor driving tasks (e.g. commercial drivers), cognitive team-based operation (e.g. air and marine traffic controllers), cognitive team-based operation combined with psychomotor skills (e.g. air and marine pilots), and one-person operation and alertness monitoring (e.g. railroad engineers, commercial drivers);
- Conduct the pilot tests within selected work environments, using simulation where appropriate;
- Investigate application of FFD/RTP testing systems to Canadian marine and rail operations;
- Investigate the use of multiple-system approaches involving in-terminal systems combined with portable in-vehicle systems;
- Conduct cross-validation studies on several FFD/RTP testing systems to examine risk factor and job performance representation (validity) for several transportation work environments.

Table 1 lists the systems recommended for pilot testing in the specified application areas.

Table 1: Systems to Consider for Pilot Testing

FFD/RTP Testing System	Application Area(s)
Readyshift®	Commercial driving
ART-90	Commercial driving
Factor 1000	Commercial driving
TOPS	Commercial driving
FIT 2000	Commercial driving Air traffic controllers Aviation pilots Railway enginemen, control operators, etc. Marine pilots, navigators
Cogscreen	Air traffic controllers Aviation pilots Railway control operators, etc. Marine pilots, navigators
NovaScan	Air traffic controllers Aviation pilots Railway control operators, etc. Marine pilots, navigators
NMRI-PAB	Air traffic controllers Aviation pilots
MABT	Air traffic controllers Aviation pilots

References

Gilliland, K. and Schlegel, R.E. (1993) Readiness-to-perform testing: A critical analysis of the concept and current practices. Federal Aviation Administration research report DOT/FAA/AM-93/13.

Miller, J. (1996) Fit for duty. *Ergonomics in Design*, 4 (2), 11-17.

Sommaire

Cette recherche avait pour but de recenser les tests d’aptitude au travail/d’aptitude à la tâche (FFD/RTP) pouvant se prêter à des applications dans le secteur des transports. Vingt-six systèmes ont été examinés. La description de la plupart des systèmes a été tirée des rapports de Gilliland & Schlegel (1993) et de Miller (1996). Plusieurs avaient aussi fait l’objet de rapports techniques. Dans certains cas, la documentation du fournisseur était la seule source d’information, difficilement accessible. Mais les chercheurs ont eu accès à toutes les données utiles concernant les systèmes les plus prometteurs.

Voici les principaux critères utilisés pour évaluer les systèmes :

- les éléments du test – les facteurs de performance évalués;
- le temps nécessaire pour passer le test au complet;
- les ressources nécessaires pour administrer le test;
- l’application antérieure du test au secteur des transports;
- le délai d’obtention des résultats;
- les études de validité effectuées sur les éléments du test, et les résultats;
- la possibilité de mettre en oeuvre l’appareillage dans un environnement de transport.

La sélection des systèmes jugés appropriés, et les recommandations pour leur mise en oeuvre, ont pris en compte les besoins suivants :

- définir un test FFD/RTP qui soit pertinent dans le secteur des transports;
- établir une théorie fonctionnelle qui devra servir de fondement aux tests;
- déterminer les buts recherchés au moyen des tests, afin d’établir le critère à mesurer et son application;
- réaliser des études de contrevalidation afin de vérifier qu’il existe bien une correspondance entre les scores aux tests FFD/RTP et le critère choisi;
- mener d’autres recherches sur les tests RTP et étudier les principes fondamentaux sur lesquels se fondent ces tests;
- déterminer dans quelle mesure la validité apparente est souhaitable, surtout si cette qualité n’est pas essentielle à un critère correspondant à un facteur de risque;
- donner une formation aux employés et aux gestionnaires pour en arriver à éliminer l’exigence de validité apparente;
- fonder les tests sur la correspondance entre le critère mesuré et les facteurs de risque ou l’exécution de la tâche, ou les deux;
- se fonder sur les résultats de la recherche et les discussions entre fournisseurs, employeurs et employés pour décider s’il faut mettre l’accent surtout sur la correspondance entre le critère et les facteurs de risque ou entre le critère et l’exécution de la tâche, ou les deux;
- s’assurer que les systèmes de test produisent des résultats fiables et qu’ils sont facilement accessibles;
- être prudent et réaliste dans la détermination des seuils de réussite aux tests RTP;
- traiter convenablement les conducteurs qui échouent aux tests, en leur accordant un suivi et un soutien adéquats;
- ne pas oublier que les tests FFD/RTP ne remplacent pas les tests de dépistage de drogues;

- se donner une période de familiarisation avec le système avant d'établir les seuils de réussite;
- faire en sorte que la durée du test soit raisonnable, compte tenu du contexte dans lequel il sera administré;
- admettre qu'il peut être nécessaire de recourir à une batterie de tests pour obtenir des résultats complets;
- s'assurer que les tests comportent des éléments pertinents pour la tâche, même si, dans l'ensemble, ils sont dépourvus de validité apparente;
- le système de test doit être peu coûteux à acheter, à exploiter et à entretenir.

Recommandations

Voici dans quelles directions devraient s'orienter les travaux futurs :

- établir le profil des emplois du secteur des transports qui ont une influence prépondérante sur la sécurité, de façon à décider quelles professions doivent être visées par des tests FFD/RTP;
- déterminer la pertinence des systèmes de tests retenus pour chaque catégorie d'emploi ainsi que le degré d'adaptation nécessaire;
- effectuer des essais pilotes de systèmes conçus expressément pour certaines catégories d'emplois, comme les tests psychomoteurs reliés aux tâches de conduite (p. ex., conducteurs de véhicules commerciaux), les tests cognitifs de travail en équipe (p. ex., contrôleurs de la circulation aérienne et maritime), les tests cognitifs de travail en équipe combinés aux tests psychomoteurs (p. ex., pilotes d'avion et pilotes de navire), et les tests qui mesurent la vigilance et la performance individuelles (p. ex., mécaniciens de locomotives, conducteurs de véhicules commerciaux);
- effectuer des essais pilotes dans les environnements de travail choisis, en recourant à la simulation au besoin;
- étudier l'application de systèmes de tests FFD/RTP aux secteurs canadiens du transport maritime et ferroviaire;
- étudier le recours à des tests multiples, soit une combinaison d'appareils fixes, implantés dans les installations terminales, et d'appareils portables, embarqués;
- mener des études de contrevalidation sur plusieurs systèmes de tests, afin de déterminer s'ils mesurent vraiment les facteurs de risque et l'exécution de la tâche (leur validité), dans plusieurs environnements de travail du secteur des transports.

Le tableau 1 présente la liste des systèmes recommandés pour des essais pilotes dans les secteurs d'application indiqués.

Tableau 1 : Systèmes recommandés en vue d'essais pilotes

Système de tests FFD/RTP	Secteur(s) d'application
Readyshift®	Conducteurs de véhicules commerciaux
ART-90	Conducteurs de véhicules commerciaux
Factor 1000	Conducteurs de véhicules commerciaux
TOPS	Conducteurs de véhicules commerciaux
FIT 2000	Conducteurs de véhicules commerciaux Contrôleurs de la circulation aérienne Pilotes d'avion Mécaniciens de locomotives, contrôleurs de la circulation ferroviaire, etc. Pilotes maritimes, navigateurs
Cogscreen	Contrôleurs de la circulation aérienne Pilotes d'avion Contrôleurs de la circulation ferroviaire, etc. Pilotes maritimes, navigateurs
NovaScan	Contrôleurs de la circulation aérienne Pilotes d'avion Contrôleurs de la circulation ferroviaire, etc. Pilotes maritimes, navigateurs
NMRI-PAB	Contrôleurs de la circulation aérienne Pilotes d'avion
MABT	Contrôleurs de la circulation aérienne Pilotes d'avion

Références

Gilliland, K. and Schlegel, R.E. (1993) Readiness-to-perform testing: A critical analysis of the concept and current practices. Federal Aviation Administration research report DOT/FAA/AM-93/13.

Miller, J. (1996) Fit for duty. *Ergonomics in Design*, 4 (2), 11-17.

Contents

Tables

1 Introduction

1.1 Background

The use of readiness to perform (RTP) or fit-for-duty (FFD) testing as an alternative to, but not a substitute for, biochemical drug testing has been considered by many safety-sensitive industries and agencies. Reasons given include better coverage of causes of poor performance such as fatigue and psycho-social stress, increased validity, and more ready acceptance by the workers. Many experimental studies (Gilliland & Schlegel, 1993, 1997; Miller, 1996) have investigated the feasibility and efficacy of these RPT testing approaches to identify those who are unfit to work. However, a need for field studies has been identified by researchers (Attwood, et al., 1994; Comer, 1995). Field validation of the operation of these systems is seen as a crucial issue at this time.

Many types of FFD/RTP testing systems exist and their strengths and weaknesses depend on the application and purpose of the tests. Computer-based cognitive tests are more relevant to the work tasks, but require expensive computer equipment and software. Other types that use some physiological measure, such as eye response, are stand-alone machines that process specific information and are easier to use. Most of these systems give immediate results upon completion of the test. However, at this time they are limited in the types of behaviour and cognitive capabilities they can measure. The basis for selection is not clear-cut, and much of the actual scientific data required to validate the tests have yet to be collected.

Perhaps an impetus for doing the research is required. It may be necessary to regulate and/or standardize the requirement for (FFD/RTP) testing, for the level of effectiveness and soundness of the theoretical basis. Regulation would force the vendors and the affected industries to do the research to validate the testing procedures. Furthermore, there is a need for the development of policy and procedures for FFD/RTP testing similar to those developed for drug and alcohol testing in the transportation industry. (See Gustafson, 1994, and Cook, et al., 1999, for details on drug and alcohol testing for transportation employees.)

Regulation would also need to determine what areas of performance apply to the various job classes, and what range of behaviour the testing should include. The inclusion of non-traditional behavioural measures as mental fitness, for instance, may be a consideration, given the recent Egypt Air accident. Such an accident raises the question as to whether the detection of a unacceptably low level of mental fitness could be determined. Of course, the measurement of mental states and their valid relationship to actual performance must be established before a useful test can be included in a FFD/RTP test.

In addition to drugs, alcohol, and stress, fatigue poses a major threat to personal and public safety in transportation. A recent safety report by the National Transportation Safety Board (NTSB) in the U.S. evaluates the efforts of government programs to address the problem of fatigue in transportation, from 1989 to 1999 (NTSB Safety Report, 1999). This report concludes that the problem of fatigue is serious, that fatigue management programs work, and that further research and development is required.

1.2 Purpose of the Present Investigation

This report is a review of:

- the research examining the scientific basis for FFD/RTP testing, particularly as it pertains to fatigue and stress;
- field and laboratory validation of the test methods;
- various FFD/RTP testing systems;
- the major issues to be considered in the application of FFD/RTP testing, particularly to transportation systems; and
- policies required in support of FFD/RTP testing.

This report provides direction for the use of FFD/RTP testing within the transportation industry in Canada.

1.3 Scope

The report consists of a literature review of all relevant materials on FFD/RPT testing, an examination of the research performed, an assessment of the testing systems, and recommendations for application of the systems to the transportation environment.

2 Literature Review

2.1 Method

The search for articles, papers, and books on RTP testing consists of:

- Extensive searches on the net using various database search engines such as:
  - PSYCHINFO;
  - MEDLINE;
  - FAA;
  - Nuclear Regulatory Commission;
  - American Petroleum Institute;
  - FHWA; and
  - NTSB.
- Journals at hand – Human Factors, 1984 to present; Applied Ergonomics, 1993 to present; HFES proceedings, 1986 to 1998; Aviation Psychology, 1991 to present; Human Factors and Ergonomics in Manufacturing, 1991 to present.
- University of Toronto's library system, including the UTLINK on-line search capability.
- Transportation Development Centre's library.
- Other libraries at various agencies, such as AECB, Defence and Civil Institute for Environmental Medicine, Ontario Hydro, and the American Petroleum Institute.
- Discussions with Ron Knipling, an expert on FFD/RTP testing systems at the Federal Highway Administration.

The relevant literature was identified and obtained for review. The quality and validity of the research were rated. The best articles and research papers were categorized as follows:

- Validation of the use of FFD/RTP testing;
- Reviews of FFD/RTP testing;
- Refutation of use of FFD/RTP testing systems;
- Discussion papers on important issues related to FFD/RTP testing; and
- Research examining the basis for FFD/RTP testing.

2.2 Selection Criteria of Research Documents

Documents were selected according to the following criteria. The document:

- focused on FFD/RTP Testing;
- contained information on testing system validity and reliability;
- dealt with the basis for choosing FFD/RTP test components; or
- discussed the reasons for the use of FFD/RTP testing systems in transportation.

3 Results of the Literature Review

3.1 Search Results

The literature search resulted in the identification of 22 papers dealing specifically with FFD or RTP testing and several others dealing with FFD/RTP issues. They represent the major contributions to the subject. It is likely that other significant works, produced by test system vendors, were not readily available. Many of the vendors did not have web sites and contact information was not generally accessible. Since the types of systems that would show good potential for adoption by the Canadian transportation industry would need to be well tested and generally available and supported, the assessment focused on those that produced accessible information. Most of these documents either reviewed specific products and approaches, and discussed the limitations of the tests (see Comer, 1995; Miller, 1996; and Gilliland & Schlegel, 1993), or provided research results that validated the test system's ability to detect risk factors or measure job performance.

A complete review of the issues and limitations of FFD/RTP testing can be found in Gilliland & Schlegel (1993). This review was completed for the FAA's Civil Aeromedical Institute (CAMI). Further specific research on the effectiveness of FFD/RTP to screen for alcohol and fatigue was carried out for CAMI by NTI, Inc. (1995). Lab research was also carried out for CAMI by Gilliland & Schlegel (1997), focusing on test learning rates and reliability of test measures. Major validation studies include Kennedy et al. (1993), Gilliland & Schlegel (1997), and NTI Incorporated (1995).

These studies have provided information useful for making policy decisions about the application of FFD/RTP testing to safety-sensitive work environments. Kennedy & Drexler (1996) examined the specificity of these tests to determine the minimum number of tests and the optimum cut-off points. Specificity identifies those who fail the test, but who are not incapacitated, resulting in unnecessary reductions in the workforce, due to false positives. This may have serious operational impact.

3.2 Description of FFD/RTP Testing Systems

Numerous FFD/RTP testing systems are available. They include those that measure some simple physiological response, a set of behavioural and psychomotor responses, both psychomotor and cognitive performance, or those that measure all of the latter, in addition to mood and sleepiness. Some are objective, while others are subjective. None at this time measures stress levels or underlying psychoses, two areas that need further investigation to identify suitable tests for inclusion into a test battery (see discussion in section 5).

The following short descriptions of the existing FFD/RTP testing systems are based on those provided by Gilliland & Schlegel (1993) and Miller (1996). Refer to these sources for more detailed information. The descriptions are given here only to illustrate the breadth of the pool of potential systems that could be applied to the transportation industry. Many have not been validated for application to transportation, however, and most would require validation studies specific to the multi-modal transportation environment.

Table 4-1 contains a comparison of these tests for pertinent factors that may affect their suitability for the transportation industry.

3.2.1 Systems Based on Physiological Response

The following FFD/RTP testing systems are physiologically based.

EPS-100

This test measures the ability of the eye to follow moving light and to react to dim and bright light. The test takes about 90 seconds to complete, with immediate results (Miller, 1996).

FIT-2000

The FIT-2000 is a FFD/RTP test system that measures the eye's reaction to light. The test measures the response of the involuntary eye reflexes, takes about 30 seconds, and provides immediate results (Miller, 1996).

Eyegaze System

The Eyegaze System determines the eye's gaze direction. The test takes a few minutes as the system records the movement of the eye, analyzing movements, fixations, and pupil dilation (Miller, 1996).

PERCLOS

This approach is an in-vehicle constant monitoring device that measures the percentage of time that the eye is closed (PERCLOS), and is being tested by the Federal Motor Coach Safety Administration in the U.S. (Dinges et al., 1998).

3.2.2 Performance-based Systems

These systems are all performance-based.

MTPB

The Multiple Test Performance Battery (MTPB) is an older system developed for the USAF Aerospace Medical Research Laboratories. It is a precursor to the NASA multi-attribute task battery (Gilliland & Sclegel, 1993).

APTS

The Automated Portable Test System is a validated, stable computerized test battery designed for the Navy (Bittner, et al., 1985). The Delta™ Readiness-to-Perform Testing system was based on this system (Gilliland & Sclegel, 1993).

ReadyShift ®

This system involves a five-minute performance test that includes a non-declining baseline of scores that account for learning and skill that develops over day-to-day and month-to-month use. The system is designed for measuring the abilities of drivers, and can be either mounted on a desktop or installed in the cab of long-haul trucks (Miller, 1996).

Delta™

The Delta™ system was developed by Essex Corporation and contains a number of mental tasks. It contains tests that produce stable results and can be learned quickly (Miller, 1996).

ART-90/Vienna Test System

This system was developed for the Austrian Institute for Road Safety for testing drivers, testing that is compulsory in Austria and Germany (Miller, 1996).

Factor 1000

The Factor 1000 system was designed and built in the 1950s and was used successfully in research on driver behaviour and fatigue (Miller, 1996). The system uses a continuous tracking task to assess driver ability.

NovaScan

NovaScan measures performance on three separate tasks (visualization, tracking, and attention) (Miller, 1996). The tests allow the assessment of skill switching. The three tasks are based on a pool of 30 tasks that are introduced randomly, three at a time. The system was developed for and adopted by the FAA to examine the possibility of testing for readiness to perform for air traffic control specialists (NTI, Inc., 1995).

Personal Safety Analyser (PSA)

The Personal Safety Analyser is a self-contained system that uses a touch-screen to allow employees to easily enter their responses (Miller, 1996). The system assesses the individual’s ability to perform representative tasks related to the workplace. The test battery includes acquisition of information, decision-making, memory, and response tasks.

DAVE

The Divided Attention Visual Experiment system was developed by Atlantis Aerospace for assessing decrements in driving performance experienced by individuals suffering from occlusive sleep apnea syndrome (Miller, 1996).

Truck Operator Proficiency System (TOPS)

The TOPS system was designed specifically for assessing driving performance and is being used by the Arizona Department of Public Safety to test truck drivers at weigh scales (Miller, 1996). If drivers fail the test they are taken out of service.

Psychomotor Vigilance Task (PVT)

The PVT system consists of a reaction-time measuring device linked to a PC computer. The test determines the level of alertness and can be used to assess fatigue (Gilliland & Schlegel, 1993).

COGSCREENTM

This system was developed for the FAA to screen pilots prior to flight. The test battery assesses the ability of the pilot to perform various cognitive tasks, a reaction time task, and tracking (Gilliland & Schlegel, 1993). The operation of the system is simple and allows input through the use of a light pen, rather than keyboard or mouse.

Criterion Task Sets (CTS)

The CTS system assesses the impact of intrusive task workload on overall cognitive performance. The greater the impact that these intrusive tasks have on degrading cognitive performance, the less acceptable the performance for continuing to work (Gilliland & Schlegel, 1993). Tests from this test battery are included in NovaScan, UTC-PAB, and STRES.

WRPAB

The Walter Reed Performance Assessment Battery is another test battery that has contributed various tests to other batteries such as UTC-PAB, AGARD-STRES, and COGSCREEN (Gilliland & Schlegel, 1993).

UTC-PAB

The Unified Tri-Service Cognitive Performance Assessment Battery was developed for the U.S. Department of Defense as a screen for drug and alcohol use and research on drug effects (Gilliland & Schlegel, 1993). This battery consists of tests taken from WRPAB, CTS, and PETER (an earlier D.O.D. system).

NMRI-PAB

The Naval Medical Research Institute Performance Assessment Battery was developed for the U.S. Navy, and has been used by the FAA in comparative studies of the impact of environmental stressors on pilot performance (Gilliland & Schlegel, 1993).

AGARD-STRES

The *Advisory Group for Aerospace Research and Development Standardized Tests for Research with Environmental Stressors* was developed to research the effects of environmental stressors on cognitive performance in pilots (Gilliland & Schlegel, 1993).

ANAM

The Automated Neuropsychological Assessment Metrics test battery uses a number of AGARD-SRES tests that are modified to screen for neurological problems (Gilliland & Schlegel, 1993).

ACS

The Assessment of Cognitive Skills Battery was developed to assess changes in cognitive states over time in physicians and other professionals (Gilliland & Schlegel, 1993). The test battery has been validated for use as a screen for cognitive impairment.

B-MAPS

Another test battery that can measure small levels of cognitive impairment is the Bexley-Maudsley Automated Psychological Screening and Category Sorting Test (Gilliland & Schlegel, 1993). This test battery has been used to assess impairment caused by the effects of alcohol, but requires validation for other risk factors.

CCAB

The Complex Cognitive Assessment Battery consists of a set of complex, high-level cognitive tests (Gilliland & Schlegel, 1993). This test battery assesses performance in various decision-making, planning, perceptual, and problem solving tasks similar to actual job-related tasks that pilots perform.

SYNWORK

The Synthetic Work Task tests for the ability to time share cognitive tasks similar to real-world situations (Gilliland & Schlegel, 1993). The test battery has a high level of face validity to the types of tasks that operators of complex systems perform.

MABT

NASA's Multi-Attribute Task Battery was also designed to test performance in complex cognitive tasks similar to those in the SYNWORK battery (Gilliland & Schlegel, 1993). The testing requires an experienced test administrator to conduct the tests. This battery is well suited to air crews, and includes tasks that assess performance on communications and resource management activities.

3.3 Research Programs

3.3.1 FAA's Civil Aeromedical Institute (CAMI)

Work at CAMI has been a major contributor of information on both the development and validation of FFD/RTP testing procedures. This work represents a long-term development project that will culminate in a set of tests to be used for determining RTP for air traffic controllers, pilots, and other personnel working at safety-sensitive jobs within aviation. So far, the results of these studies (Gilliland & Schlegel, 1993; 1995; 1997; and NTI Incorporated, 1995) have been somewhat inconclusive, but have highlighted the major problems of applying FFD/RTP testing procedures within the aviation environment, and the need to validate the tests for each application (Gilliland & Schelgel, 1995).

3.3.2 Other Validation Studies

Validation studies by Kennedy et al. (1993) of FFD/RTP methods used for detecting the effects of alcohol on various mental performance tests, has shown how well these test batteries can identify individuals having a blood alcohol concentrations (BAC) that may pose a risk to job performance. These researchers also conducted experiments to determine the rates of false positives (indication that a person is "over-the-limit", when in fact they were not) (Kennedy et al., 1995) and what the cut-off values were for certain test battery components (Kennedy et al., 1993; 1994; and 1995).

3.4 Significant Documents

The works of Gilliland & Schlegel (1992; 1993; and 1995) are considered the most comprehensive reviews of the theory and application of RTP test methodology. A more informal review of the various test methods has been presented by both Miller (1996) and Comer (1995). Various documents on the impact of mood, personality, and stress on performance were also consulted (Parkinson et. al., 1996; Hockey, 1986; and Stokes & Kite, 1994).

3.5 Areas of Focus for Test Components

The main areas of focus for the components assessed by any "ideal" FFD/RTP testing could include the following:

- Physiological response
- Reaction time performance
- Cognitive performance
- Subjective mood
- Measurement of stress levels
- Personality profile

Most of the performance-based testing systems include the second and third components, reaction time, and cognitive performance. Some include, in addition to these two, the fourth component, subjective mood. Physiology-based test systems test a specific response, such as eye movement or pupil dilation. The last two components do not exist in any of the test systems identified here. This study will investigate some tests that could be included, but further study and validation will be required to determine their efficacy as assessments of risk factors or potential job performance.

Combining physiological tests with cognitive ones may be a solution to the problem of trying to fit more than one test session in a single workday. The physiological tests could be performed at some other time during the work shift, such as during a break.

4 Major Issues

Gilliland & Schlegel (1993) made a series of recommendations in their review of FFD/RTP testing systems. It is instructive to list those recommendations here, because they point to some of the major issues that have been raised about FFD/RTP testing by others as well.

- There is a **need for a clear definition** of what a specific FFD/RTP test is within the context of the application – i.e. within a particular working environment.

- **A working theory** for which the FFD/RTP testing must be based, needs to be established before validity in various work environments can be accepted.

- A **clear idea of the objective** of the FFD/RTP must be determined to establish the measurement criterion and the use of that measure – i.e. for prediction of either job performance, and/or the risk factors that may potentially affect job performance.

- **Cross validation** studies must be carried out to ensure that the FFD/RTP tests do in fact show a true relationship (predictive) to the criterion chosen.

- **Further research** (actually very little exists even now) needs to be done to both assess the RTP test methods and to explore the fundamental principles on which the testing is based.

- There is a need to determine whether **high face validity** is desirable, particularly if that validity is not essential to a test criterion that predicts risk factors.

- Can **education of the employees and management** eliminate the need for face validity?

- Tests can be based on criterion **validity for either risk factors or work performance, or both**.

- **Research results** must be used to make decisions on whether to focus on risk factor criterion validity and/or for work performance criterion validity.

- Care and realistic expectations must be used to determine the **"cut-off" values** for RTP tests – i.e. the failure point should not be based solely on the number of standard deviations from the mean value.

- **RTP testing is not a substitute or replacement for drug testing** (see Kennedy et. al., 1995; 1996, for the limitations of RTP testing).

- Test duration should be within **practical limits** given the operational environment.

- **Multiple tests** may be needed for proper coverage.

- Test systems must contain elements that are **relevant to the job**.
- The testing must be **affordable** to purchase, operate, and maintain.

4.1 Definition of the Test's Purpose

The test's purpose in the context of the job must be clearly defined. If the purpose of the test is to assess some physiological or performance measure that is highly correlated to a certain level of intoxication or fatigue, then this must be stated and adhered to during any decision-making. Alternatively, the testing may be measuring actual job-related performance used as a set criterion. This clarification will make the limitations of the test apparent and will allow the workplace evaluator/test administrator to make an informed decision as to whether the subject of the test can work at the assigned job or not. The evaluator/test administrator can weigh the results against other criteria based on experience and knowledge of the subject.

4.2 Working Theory

The test must be based on a working theory that provides the structure for validity tests, the administration of the test, and interpretation of results. This theory should be borne out in independent research results, and should make sense in the context of the work environment. If the FFD/RTP testing is to be applied to a trucking environment, then the theoretical basis for the testing should consider factors that are germane to the commercial driver's performance, such as reaction time, vigilance, logical-reasoning, and spatial orientation. The test results and cut-off limits should reflect the expected performance levels deemed acceptable, based on a theoretical model that considers all of the variables and their relationships. The model must be dynamic and predictive, although it may be based on empirical data (Gilliland & Schlegel, 1993).

4.3 Test Goals

The goals of the testing must be established and clearly stated. Is the FFD/RTP testing designed to determine whether an employee is incapacitated from drugs, alcohol, fatigue, or stress, or does the testing actually measure performance that is job related? Or is the testing able to both detect drug, alcohol, fatigue, or stress induced behaviour, and measure job-related performance.

The levels of acceptance, or cut-off points for the criterion must be clear and equitable. If the testing involves more than one test, the values for acceptance for each test and their weightings must be considered according to a specific method (statistical or mathematical). All conditions for interpretation of results must be stated and have recognizable criteria. The test method should clearly state whether the cut-off points are based on the individual's own baseline results established from a set of previous test trials, or based on some standard determined from the research literature, data collected by the vendor, and pilot testing at the site where the testing is to take place.

4.4 Test Validity

The acceptance of and trust in the FFD/RTP testing methods will be partly a function of the validity of the testing, in addition to other factors such as ease of administration and support from immediate and upper management. However, validity of the test with respect to reflecting job performance will be a major contributor to the widespread acceptance by employees of its worth and value to safety and their expected performance. The positive impact of the testing on improvement in worker and public safety should be demonstrated by past empirical research or be seen in initial on-site pilot studies conducted prior to adoption. Such research or pilot studies should consider the following factors, on the job:

- Accident rates;
- Reported errors;
- Performance; and
- Reports of irregularities.

Two types of validity must be used to determine the appropriateness of a FFD/RTP testing method for application to a specific work environment:

- Criterion validity; and
- Face validity.

The former refers to the ability of the testing to predict the behaviour it sets out to measure. The latter is how well a test appears to measure the risk factors or job performance. Acceptance of the use of criterion validity may require some education of management and personnel to fully appreciate the strengths and relevance of the criterion.

4.4.1 Criterion Validity

It is important to ensure that the tests included in the FFD/RTP testing have been shown to reliably measure either risk factors or performance. That is, the tests must be validated through careful studies to determine how well the tests detect the risk factors (effects of alcohol, drugs, fatigue, stress, etc.) or predicted job performance. Results from tests collected during studies using other similar tests cannot reliably validate a specific set of tests or test battery.

4.4.2 Face Validity

The issue of using face validity in FFD/RTP testing is that many test methods rely on face validity to garner acceptance by employees and employers. If the tests appear to be measuring attributes that are similar to job performance tasks, an impression is formed that this is the same as measuring actual job performance. The danger here is that criterion validity may be diluted in order to increase face validity. If face validity is important and must be emphasized, the tests must show consistent criterion validity as well.

4.4.3 Impact of Face Validity on Criterion Validity

Gilliland & Schlegel (1995) emphasize the care that is necessary when adopting tests for use in the field and when making modifications to tests that are already validated in the lab. They point out that, from their previous research (Gilliland & Schlegel, 1993), very minor changes such as changes in "inconsequential formatting features" to a test, can result in the test losing its validity as a predictor of either risk factors or job performance. Hence, modifying a test to increase its face validity, to improve its acceptance by personnel (i.e. it looks more like their job tasks), may reduce the test's criterion validity. This may then render the test ineffective at predicting risk factors, or even related job performance (i.e. certain cognitive or behavioural tests may be good predictors of job performance, but may not even resemble job tasks).

4.5 Test Reliability

4.5.1 Factors Affecting Test Reliability

Internal Reliability

The test battery must have inherent consistency in test results such that during practice trials similar results will occur from trial to trial when conducted under near identical conditions. That is, if the same person is tested when rested and sober, the results would not vary more than, for example, one standard deviation. The same must hold if the individual has had exactly the same amount of blood alcohol, components of a drug, level of fatigue (e.g. measured using a multiple sleep latency test), or level of stress (e.g. as measured by comparative heart rate and blood pressure). Kennedy et. al. (1995; 1996) found a significant number of false positives (indications of inebriation when in fact the individual was sober and well-rested) in an FFD/RTP testing procedure their research was validating. This can cause problems with acceptance and support for the testing. Their research shows that it is difficult to prevent the tests from resulting in a significant number of false positives.

External Factors

Since the impact of external factors can influence test performance, the validation tests must be conducted under conditions similar to those found in the FFD/RTP testing environment. Hence, control over external variables during the FFD/RTP testing sessions should be considered, such as administering the test in a separate room with little distractions or stress-inducing stimuli. However, if such controlled conditions are near impossible to achieve in the workplace, simulation of those distractions common to the work environment will need to be included in the validation study.

Impact of Practice

The impact of practice on a test's stability and reliability cannot be overemphasized. The amount of practice and refresher training will directly affect the reliability of the tests so that long spans of time between testing sessions may reduce performance and possibly result in a failure to pass the test. The level of practice prior to establishing a baseline will affect the reliability of the use of cut-off points. Too little practice and the established baseline will fall far short of the individual's ability to perform the test battery at a later date, while sober and rested, but may also be less than that individual's performance while incapacitated. Furthermore, the more times the individual performs the test battery, some learning and skill development may still occur. After some time the same individual may be able to easily pass the test even while incapacitated by fatigue, alcohol, or drugs. The selection of test components must take into account the amount of practice that is required for the test to become stable and less variable under similar circumstances (i.e. when sober and rested, and not stressed). If a test does not show a leveling out of the learning effect within a reasonable number of practice sessions, it may be wise to drop the test and look for one that will.

4.5.2 Issues Arising from the Level of Test Reliability

Acceptance of the Tests

If the FFD/RTP test battery is not consistent in its ability to either detect risk factors such as inebriation, fatigue, or stress, its acceptance by both management and personnel will be compromised. Just a few failures to detect these risk factors will result in distrust of the system and the entire FFD/RTP test program. Some experimental programs have recorded occasions where individuals who are over the legal limit for blood alcohol content have passed the FFD/RTP test battery (see Kennedy & Drexler, 1996), as well as in other anecdotal reports (Comer, 1995). Such results usually weaken the acceptance of the testing as a trustworthy indication that a person can or cannot work. Employees will not feel that the testing is a fair test of their capability to work, and will not support the testing program. This could likely prompt union grievances, work slowdowns, or other similar reactions.

Other reliability problems such as inconsistencies in results due to equipment calibration issues, or poor equipment availability, will compromise the company's ability to get acceptance from workers and can also result in frustrations for the test coordinator. Support for the testing will be eroded and the program is likely to fail.

Legal Implications

The reliability and validity of FFD/RTP testing systems available today have not been accepted by the legal profession as a credible indication of risk factors or job performance (Comer, 1995). In fact, biochemical-based drug testing has only recently been accepted officially by the courts (Gustafson, 1994). Since there is no hard evidence of the risk factors (drug, alcohol, fatigue, stress, etc.), the FFD/RTP testing will be an even more difficult sell to the legal system. On the other hand, Miller et al. (1994) report that drivers who failed the Arizona Department of Public Safety's FFD/RTP test administered at weigh stations were mostly accepting of the results. None of these cases ever reached the appellate court level. Another issue is the use of the self-referenced baseline. If a worker who falls below his/her own baseline, and consequently fails the test, but is still performing higher than fellow workers with lower baselines, may have a legitimate grievance or grounds for litigation.

4.6 Effect of Failing the Test

Many researchers and authorities in the field of FFD/RTP testing concur that the response of the workplace to those who fail the test will be the most difficult hurdle to clear (see Gilliland & Schlegel, 1993; and Comer, 1995). There is little evidence that failing an FFD/RTP test session may result in the same kind of stigma attached to failing a biochemical-based drug test. However, some anecdotal information from previous experiments and vendor claims do support this notion (Gilliland & Schlegel, 1995). Furthermore, employers are concerned that if an employee fails a test session, that employee’s absence will increase the workload of the employees left to handle the shift. Calling in someone to replace the unavailable employee will not be feasible where shiftwork is the norm, since the probability of finding a well-rested replacement to work a midnight shift, for example, may be quite low.

4.7 Test Criterion

The test criterion must be determined carefully, and stated clearly so that all parties are cognizant of the purpose of the testing, i.e. is the testing designed to detect the risk factor (impact of fatigue, alcohol, or drug effects on certain mental functioning) or the potential impact of these risk factors on actual job performance? Ross and Mundt (1996) provide valuable insight into the problems that are common to the measurement of actual job performance, such as high individual differences (tolerance levels and performance), compensatory behaviours, and the nature of tasks (number, complexity, immediacy, and novelty). These issues make test battery development difficult and provide the acid test for suitability for high job performance validity. To date very few test batteries meet this expectation of high job performance validity (refer to Miller, 1996, for a review of these problems as they apply to the transportation industry). The test criterion must consider the:

- ability to measure alcohol, drugs, fatigue, emotion, stress, mental state;
- performance measures used;
- basis for the establishment of cut-off values; and
- use of a self-referenced baseline versus the use of a lower limit based on some determined acceptable level.

4.8 System Availability

The system must always be available during the period in which employees must be tested. Breakdowns, either hardware or software related cannot be tolerated. Personnel will be suspicious of the results if equipment is not working properly, or seems to be in poor repair. If the system is unstable or tenuous, it is not likely to be taken seriously and the whole testing program will be jeopardized.

4.9 Applicability of the Field and Lab Research

4.9.1 Theoretical Basis for FFD/RTP Testing

A sound theoretical basis for using certain tests for FFD/RTP methods must be established through laboratory testing. The actual criterion validity of each test component must be tested and its ability to detect risk factors assessed. This involves both sensitivity to the risk factor and specificity, as well as other factors such as reliability, stability, usability, and other forms of validity (Kennedy & Drexler, 1996). Specificity involves the percentage of participants who fail (fall below their baseline) who were not affected by a risk factor (false positive). False positives can be as much of a problem as can false negatives (the test misses a participant who is affected by a risk factor).

The fundamental construct of the test (relationship to cognitive functioning or reaction time, for example) will be an important factor in making the decision about how much to adjust the test cut-off points up or down. Also, if a certain test involves cognitive or psychomotor elements that map directly into important job task elements, cross validation will be more likely.

4.9.2 Cross Validation Research

Many studies have been conducted to determine the usefulness of various cognitive and psychomotor performance tests to detect a risk factor such as alcohol, drugs, fatigue, or stress. However, such studies cannot be directly related to job performance in a particular work environment. Gilliland & Schlegel (1993) recommend that cross validation studies be performed to test the relationship between FFD/RTP test battery performance and that of specific job task performance. A simulated work environment could be used to conduct a double-blind experiment where neither the participants in the study, nor the researchers are aware of who has been given a drug, alcohol, or has been sleep deprived. Such a study could look at specific test scores for individuals and correlate those to specific job tasks that can be measured, such as errors made, response time, or ineffective task completion.

4.9.3 Field Studies

Comer (1995) carried out a review of several field-based and laboratory studies. The author points out the need to test the systems in a field setting. This may involve some type of simulation or pilot test. The study would not be as controlled as a laboratory study, but would “test” the ability of the FFD/RTP system to function in a real setting. Such field tests should follow lab-based cross validation studies, as described by Gilliland & Schlegel (1993). Unfortunately, data collected by vendors on the effectiveness of their FFD/RTP systems to detect risk factors or their relationship to job performance, are not generally shared with the scientific community. Hence, a wealth of empirical data may exist, but not be available for validation purposes.

4.10 Purchase and Operating Costs of Test Systems

The initial cost of the FFD/RTP system is usually considerably lower than the equipment required to do drug testing. However, administration of the tests will cost more if the FFD/RTP testing is to be done on a daily basis (more often than drug testing), a recommendation by most vendors and researchers. On the other hand, drug testing may be less effective because of the spottiness of the testing and the lack of daily monitoring. Other costs that affect drug testing include the cost of requiring staff to cover for personnel who fail the testing, and for the considerable costs of follow-up and counseling. However, these are important parts of a successful testing program. The costs and benefits must be weighed ahead of the decision to implement such a program.

4.11 Applicability of the Tests to Transportation Systems

4.11.1 Comparison of Existing FFD/RTP Testing Systems

The systems shown in Table 4-1 consist of two basic types: those that use some psychophysiological measure, such as eye movement or response, and those that use a battery of cognitive/psychomotor tests. The following comparison considers these variables:

- The test system components – what performance factors the test evaluates;
- The time required to complete the entire test;
- The resources required to conduct the testing;
- Past application to a transportation environment;
- Turnaround time – when test results are available;
- Validity studies that were done on the test components and measures; and
- A rating for the feasibility of applying the device to a typical transportation work environment.

Systems that were considered were those that were close to ready-to-use or already in use by transportation operators. Those systems still in development requiring lengthy test times or analysis, or not containing test components appropriate to the transportation environment, were not pursued. The emphasis was on practical application to operations in transportation systems. If a testing system could not be readily installed, set up, and used within a few days, it was considered to be still developmental. However, these systems may be viable in the future, and may introduce improvements.

4.12 Summary of Results from Table 4-1

4.12.1 Rationale for Column Headings and Rating Scheme

Column Headings

System Name: The proper name of the technology is represented by the acronyms listed in Table 4-1 and defined in section 3.2.

Test Components: The types of tests are listed for each system. These test components make up the whole battery of tests provided, but actual testing sessions may only contain a subset of the components.

Time to Complete: This column gives the completion time for each test session, where such information is available.

Resources Required: The resources required to run the test sessions are listed, including equipment and personnel. Where the equipment is wholly provided as a single hardware/software unit, requiring no set-up (not including occasional recalibration), the word "self-contained" is used.

Past Application to Transportation: This column contains references to past usage in transportation studies and actual application to an operational transportation environment.

Turnaround: This column refers to the amount of time it takes to provide results of the test session.

Validity Studies: The papers listed here represent major research done to validate the FFD/RTP testing system in a lab or field setting. The validation may be of the test criterion itself, or the use of the technology in a specific operational environment.

Rating: The rating is a scale that represents the level at which the FFD/RTP testing system may fulfil the requirements of successful application to the transportation environment. The rating is based on:

- its past application to the transportation industry;
- the amount of third-party validity data available;
- the duration of the test;
- the availability of the results (immediate versus longer duration due to analytical requirements);
- the potential reliability of the system (i.e. a self-contained, field proven system versus lab-based collection of equipment that is still experimental); and
- the appropriateness of the actual test elements.

The last variable is based on the theoretical construct of the tests included in the test battery.

4.12.2 Rationale for Application to Transport Operations

The FFD/RTP testing systems were considered applicable to transportation if they achieved a "1" in the rating. However, their specific application will be determined based on appropriateness to a particular transport operation. For example, systems that include tracking, simple decision-making, and perception tasks would be suited to commercial driving. Tests consisting of complex cognitive tasks would be more suited to air traffic control, piloting aircraft, or piloting a ship. Physiological tests could be used as a subsequent check for negative effects of fatigue or a performance degrading substance.

A number of systems exist that are well suited to, and have been validated for, driving tasks. In fact, some are in use by various transportation authorities, such as:

- *Factor 1000*, adopted by the FAA for trials to use in monitoring commercial drivers;
- *TOPS*, used by the Arizona Department of Public Safety at weigh scale stations to monitor commercial drivers; and
- *ART-90*, a test system adopted by the German and Austrian departments of highways.

Some systems have been validated but have a high probability of adoption for use by a transportation operation including:

- *PERCLOS* has been validated by the FHWA for in-cab on-going detection of fatigue during commercial driving operations;
- *FIT 2000* is under review by the Federal Motor Coach Safety Administration (formerly the FHWA OMC); and
- *Readyshift*® has been validated for use in over-the-road operations.

At present, only four FFD/RTP testing systems have been validated for more complex tasks:

- *NovaScan* has been validated by the FAA for air traffic controllers; and
- *NMRI-PAB*, *COGSCREEN*™, and *MABT* have been validated for pilots.

Many of these systems require further work to make them more self-contained and robust. They include *NovaScan*, *NMRI-PAB*, *COGSCREEN*, and *MABT*, all of which require analysis to determine fitness for work and consist of PC computers and associated equipment involving special set-up. The step to automate the analysis and consolidate equipment into a single unit is available and should be considered.

Some promising systems still require validation:

- *DAVE*, for commercial driving applications;
- *EPS-100* and *EYEGAZE*, for subsequent and in-cab monitoring;
- *CTS, MTPB, WRPAB, AGARD-STRESS*, and *UTC-PAB*, for complex operations.

Many of these systems, however, have a great deal of overlap with those that are already validated. Hence, it seems more reasonable to adopt those closer to readiness.

Table 4-1: Factors That Are Important for Assessing Suitability of Test Systems for Application to Transportation Industry

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating*
MTPB	• Monitoring • Arithmetic • Complex Code-solving	N/A	PC (FAA dev'd) Requires analyst	FAA – CAMI	Analysis	Chiles, et al., 1968; 1972.	2
APTS	• Monitoring	N/A	PC (USN)	None	Analysis	Kennedy et al., 1985 Merkle et al., 1985	3
ReadyShift®	• Monitoring • Tracking • Reaction time	5-10 min.	Self-contained	Driving tasks	Immediate	Miller et al., 1994	1
Delta™	• Linguistic • Memory • Spatial • Perception • Reaction time	N/A	PC Requires analyst	None	Analysis	Based on APTS	3
ART-90	• Monitoring • Tracking	N/A	Self-contained	Driver testing in Germany and Austria	N/A	Institute of Road Safety in Austria	1
Factor 1000	• Tracking and reaction time	2 min.	Self-contained	National Highway Transportation Safety Administration, and the Federal Highway Administration	Immediate	NHTSA and FHWA validated	1

* The rating is based upon whether the system contains appropriate test elements, validation or assessment within a transportation environment, amount of time to complete the test, immediacy of the results, and required equipment. "1" represents a system that meets all criteria.

Table 4-1 continued

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating
NovaScan	• Visualisation • Tracking • Monitoring • Attention • Logical reasoning • Decision-making • Mathematical • Memory • Situation awareness	3-10 min.	PC Special input device Requires analyst	Air traffic control – FAA	Analysis	O'Donnell (1992)	1
PSA	• Information encoding • Decision-making • Long-term memory • Response selection and execution	N/A	Self-contained	N/A	N/A	N/A	3
DAVE	• Tracking • Reaction time	N/A	Self-contained	N/A	N/A	N/A	3
TOPS	• Hand-eye co-ordination • Attention	8 min.	Self-contained truck-cab simulator	State of Arizona Dept. of Public Safety	Immediate	N/A	1
PVT	• Reaction time – unalerted	10 min.	PC or a self-contained hand-held box	NASA and FAA project for pilots	Analysis	Graeber et al. (1990)	1

Table 4-1 continued

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating
CTS	• Display Monitoring • Unstable tracking • Interval production • Continuous recognition • Grammatical reasoning • Linguistic processing • Mathematical processing • Memory search • Spatial reasoning	Variable depending on number of tasks and task variation	PC Requires analyst	USAF Aerospace Medical Research Laboratory Military pilots	Analysis	Schlegel and Gilliland, 1990	2
WRPAB	Subset of: • Logical reasoning • Reaction time • Spatial reasoning • Pattern matching • Mood • Sleepiness • Mathematical processing • Spatial judgement • Colour & form discrimination	10 to 20 min.	PC and timing board Requires analyst	NavCanada and Transport Canada for air traffic controllers	Analysis	Thorne, Genser, Sing, and Hegge, 1985	2

Table 4-1 continued

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating
UTC-PAB	Combination of WRPAB, CTS, and PETER tests	30 – 40 min.	PC and timing board Requires analyst	Developed for the U.S. DOD	Analysis	Hegge, Reeves, Poole, and Thorne, 1985	2
NMRI-PAB	Subset of: • Pattern matching • Colour & form discrimination • Logical reasoning • Reaction time • Short-term memory • Attention • Response accuracy • Response acquisition • Spatial orientation	30 min.	PC and timing board Requires analyst	Developed for the U.S. Navy FAA evaluated validity for pilots	Analysis	Schrot and Thomas, 1988 Horst and Kay, 1988	1
AGARD-STRES	Subset of UTC-PAB	30 min.	PC and timing board Requires analyst	For AGARD to assess impact of environmental stress	Analysis	Schlegel and Gilliland, 1992	2
ANAM	Subset of AGARD-STRES	20 min.	PC and timing board	Neuro-psychological screening	Analysis	Schlegel and Gilliland, 1992	2
ACS	Thirteen tests – • Linguistic • Decision-making • Memory Other types not specified	N/A	PC Requires analyst	Developed for health care workers	Analysis	Powell, Catlin, Funkenstein, Kaplan, Ware, Weintraub, and Whitla, 1990	3

Table 4-1 continued

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating
COGSCREEN™	• Tracking • Memory • Attention • Reasoning • Spatial perception • Reaction time	N/A	Self-contained PC-based system with touch screen and light pen Requires analyst	FAA pilot screening	Analysis	Horst and Kay, 1991	1
CCAB	Subset of: • Attention to detail • Perception • Time-sharing • Comprehension • Verbal reasoning • Quantitative analysis • Planning • Decision-making • Situational assessment • Problem solving • Creativity • Memory	N/A	PC	None	N/A	Geiselman and Samet, 1986; Kay and Horst, 1988	3
B-MAPS	• Decision-making • Spatial reasoning Other types not specified	45 min.	PC	Automated test used to assess alcoholics	Immediate	Acker and Acker, 1982; Glenn and Parsons, 1990; 1991	3

Table 4-1 continued

System Name	Test Components	Time to Complete	Resources Required	Past Application to Transportation	Turn-Around	Validity Studies	Rating
SYNWORK	Presented simultaneously • Memory • Arithmetic • Visual monitoring • Auditory monitoring	N/A	PC	Application is not available	Immediate	Kane and Kay, 1992	3
MABT	• Tracking • Resource management • Response time • Visual monitoring • Auditory communication • Subjective workload	N/A	PC	Developed at NASA Langley Research Centre for assessing aircrew operator performance	N/A	Arnegard, 1990; 1991	1
EPS-100	• Light tracking by eye • Eye's reaction to dim and bright light	90 sec.	Self-contained unit	N/A	Immediate	N/A	2
FIT-2000	• Pupillometric • Nystagmus	30 sec.	Self-contained unit	Under review by FMCSA (formerly the FHWA OMC)	Immediate	Rowland et al., 1997; Russo et al., 1999	1
EYEGAZE	• Eye's gaze direction	2 min.	Self-contained unit	N/A	Immediate	N/A	2
PERCLOS	• Measures the percent of slow eye closure	constant	Self-contained eye monitor	FMCSA is testing the system for commercial drivers and pilots	On-going analysis	Dinges et al., 1998 Wierwille et al., 1994	2

4.13 Resources Required for System Operation

The costs of the various systems on the market vary, but are within a range that is consistent with most computer systems and associated software licences. The cost of government developed systems may be somewhat less, but may also require some work to make them more usable in an actual testing environment. The FFD/RTP testing system vendors have considered the need for robustness and usability as key ingredients for successful implementation. Some systems may require highly trained professionals to interpret results or administer the tests, while others are stand alone self-contained systems that cannot be tampered with. The latter type of system is more appropriate for a work environment.

4.14 Test Flexibility (Customization)

Some work environments will require customization of the types of tests used in the test battery. In fact, Gilliland & Schlegel (1993) emphasise the need to provide a selection of tests that can be representative, and that can be randomly presented to reduce boredom and continued learning. The alternative is to continually establish baselines, if the cut-off points are based on these, or continual adjustments to an across-the-board cut-off point, if this type of approach is used.

4.15 Test Frequency

It must be decided how many times during a shift, personnel should be tested. Some researchers feel that the best approach is to test at the beginning of the shift and during the mid-shift break (Gilliland & Schlegel, 1993). However, this may not be feasible in many transportation environments where the shifts are irregular and personnel are on active alert throughout the entire shift (e.g. pilots, air traffic controllers, ice navigator). Portable equipment that involves a short (less than 5 minutes) test where results are available immediately, may be an ideal system for these kinds of work conditions. Examples of such systems include those that are physiologically based such as EPS-100, FIT-2000, and Eyegaze. Certainly, once per shift should be the minimum if testing is to be effective.

4.16 Implementation and Employee Acceptance

4.16.1 Ethics

The reported sentiment in the literature is that any form of testing, biochemical or otherwise, will result in some form of invasion of privacy (Comer, 1995; Gilliland & Schlegel, 1993; Miller, 1996). However, most agree that FFD/RTP testing is less intrusive and does not infringe on a person’s private life outside of work. For this very reason, it is likely that FFD/RTP testing will be accepted more readily than drug testing.

The way in which the company and employees view a failure to pass the tests will influence how successful rehabilitation programs will be and acceptance of the testing by the union and the employees. Other ethical issues arise when the pass/fail criterion is considered (see section 4.16.2).

4.16.2 Pass/Fail Criterion

What is the pass/fail criterion? Is it a cut-off point based on the individual's own baseline? Or is it an across-the-board level agreed upon by the employer and employees? If it is based on a baseline value, is there a chance that the individual managed to do poorly enough to allow him/her to pass even when affected by some risk factor? What proof is there that such a scenario cannot happen? What if one person's baseline is much higher (better) than another's, and he/she fails, but still has a score higher than the second person (with a lower baseline), and yet that second person can still work? Is this equitable? These questions raise ethical issues and will be difficult to resolve. The company, employees, union, regulatory agencies, and the FFD/RTP testing system's vendor, all must work hard to find some common ground and agree to a set of rules and procedures (see section 4.12.3). Education on the basis for the test criterion and individual differences may help to solve these problems.

4.16.3 Involvement of Employees, Union, Company, and Vendor in the Process

The initial development of the FFD/RTP Testing program should involve all of the stakeholders. Failure to do so will result in sabotage of the program in many areas, and any goals set will not be met. Resources will be wasted and any further attempts at implementing other safety programs will be hampered.

5 Application Issues for the Transportation Industry

The following issues should be considered when developing or adopting a FFD/RTP testing program and implementation plan. Each of these issues will have consequences for success of the program. Some of these issues have been encountered by organizations trying to set up and run drug-testing programs that assess biochemical agents present in the body. This pharmacological approach has been used in many work environments and a considerable amount of work has been done to validate the tests and programs (see Potter & Orfali, 1990; and Combs & West, 1991). The Department of Transportation in the U.S. has enacted a law that forces transportation agencies, companies, and organizations to set up, administer, and maintain drug and alcohol testing procedures. These procedures are detailed in Gustafson (1994).

Application Issues

- Setting for the testing;
- Portability of systems;
- Frequency of testing;
- Response to a failure;
- Professional licensing issues;
- Union concerns;
- Criterion vs. face validity for tests applied to transportation;
- Other supporting channels such as EAP and training;
- Past work in fatigue and fatigue management;
- Best type of FFD/RTP testing system for transportation areas;
- Inclusion of stress assessment and personality screening;
- Mode-specific versus multi-modal application; and
- Organizational policies.

5.1 Practical Matters

5.1.1 Setting

The testing must occur in a separate private area free of distractions, private, and preferably closed off from the day-to-day hustle and bustle, but easily accessible as employees enter their work area. The tests take some time, up to 15 minutes or more if a second test must be performed (if the employee fails the first one). Therefore, the ideal area is near the employee entrance or specific work area (e.g. entrance to a control room, employee flight preparation room, operations area).

5.1.2 Portability

The system should be portable enough to relocate easily if used in more than one area, or if it must be set up in a temporary space, and then stored. Permanent systems should not be used for settings where frequent movement of the system is expected. The ability of the system to remain available and functioning properly is a key factor in its acceptance and longer term viability.

5.1.3 Frequency of Testing

The decision as to how often employees must be tested is still area of discussion. Tests based on performance factors can provide immediate feedback, unlike those using pharmacological methods (see Comer, 1995). Hence, it is feasible to test employees prior to each shift, providing that an adequate amount of time can be fitted into the employee's normal workday. Also, it has been shown that employees become unpracticed in their performance of the FFD/RTP tests if they leave some period of time (typically just a few days) between test sessions. This may result in the employee's failure of a test even when he/she is not inebriated, fatigued, or stressed (Gilliland & Schlegel, 1993).

5.1.4 Test Duration

The amount of time it takes for employees to take the test must be short (between 5 and 10 minutes). This is evident from field-based research (Miller et al., 1994) and is a result of the constraints of the work environment. It is at additional cost that FFD/RTP testing be incorporated into the daily routine. The time factor is the most important, since schedules are tight and the transportation employee will often already be working overtime. Hence, tests of long duration will be impractical.

5.2 Response to FFD/RTP Test Failures

The transportation industry must develop policies, strategies, and procedures to support employees who fail a test. There must be programs in place that can place the employee, temporarily, into a job that is not safety sensitive, or be able to send the employee home, and have someone else (e.g. a supervisor) step in to take over the responsibilities. Furthermore, the employee who fails a test must be provided with the best information and counseling to help identify any problems with substance abuse, excessive fatigue, or unduly stressful life or job situations, and find the best course of action to reduce and/or eliminate the problem.

Anything less than this will likely lead to failure of the FFD/RTP testing program. In fact, it is probable that unions would look at inaction as a clear abuse of the testing program, and litigation may follow. Also, education about the program and available literature on its effectiveness in other environments, would help to alleviate any misunderstandings and garner support (Comer, 1995).

5.3 Acceptance by Unions and Employees

The aviation, rail, and trucking industries have long histories of strong union support for employees. Test validity, purpose, and reliability must be clearly stated and demonstrated. The whole program must be presented and explained such that both employee and employer representatives are in agreement with program procedures and goals. A comprehensive educational component will also be necessary to help ensure employee acceptance. Considerable experience and information are available on drug testing programs that have many similarities and parallels to FFD/RTP testing programs. In fact, much of the literature on drug testing stresses that education, provision of valid, reliable tests, and effective follow-up (e.g. employee assistance programs) are the keys to success (see Combs & West, 1991; Potter & Orfali, 1990).

Many unions see drug testing, for instance, as an infringement on workers' rights, and are suspicious of the employer's reasons for implementing the programs (Canadian Centre for Occupational Health and Safety, 1987). Unions point out that it is a fallacy that a drug-testing program will reduce accidents and improve job efficiency and productivity. They explain that most accidents and hazardous conditions are a result of many other factors, such as working conditions, training, and job satisfaction.

On the other hand, unions do recognize that there are steps to be taken to ensure safe working conditions and improved public safety. These steps involve the responsibility of both the employer and employees. Employers must ensure a safe working environment, and must take ultimate responsibility for public safety. The employee must do all that is reasonable to ensure that he/she is rested, free of alcohol or drugs, and mentally and physically fit to perform his/her duties as required by the collective agreement, if the employee is a union member, or by a work contract between the employer and employee.

The purpose of FFD/RTP testing is to help ensure that employees can meet their obligation. However, can such testing procedures actually do this? Does this type of testing actually reduce hazardous conditions, improve job performance, and increase public safety? As shown above, enough work has been done to show that fatigue, stress, alcohol, and drugs often do degrade performance. However, it is a leap to then deduce that screening out those who may be under the influence of alcohol or drugs, or who may be very fatigued or stressed, will improve safety or job performance (see Gilliland & Schlegel, 1993, for more discussion).

The legal standpoint of the union is that unless the company can prove that an individual is "not fit to work" within some set of agreed upon guidelines, the employee cannot be kept from working. The testing must be able to show without doubt that the person being tested is in fact "not fit for work".

Hence, validation studies must be conducted for the specific test battery, in a hi-fidelity simulated work environment, that allows the measurement of job performance. Only with this information will a union be obligated to consider the utility of FFD/RTP testing as a responsible and equitable way of screening personnel prior to working.

5.4 Follow-up and Education

It is clear from the information on the effectiveness of drug testing programs, that testing alone is not enough. In fact, the program may suffer without some form of follow-up that supports the stated purpose. Such a situation may place the organization administering the test in a very poor legal position. The test may very well be viewed as a witch hunt, with no real positive value to the employees. Hence, the employer must be committed to ensuring that if an employee fails the test, the employee will have the opportunity to learn how to effectively control his/her problem. This assistance would likely be in the form of an employee assistance program (EAP) that provides the professional help and education necessary to help employees control their risk of substance abuse. Such programs involve counseling, training, and self-reports of problems that may have led to the use of substances or fatigue.

Training and education should be provided to all employees to make them aware of the risks and how they may better cope with the conditions that may cause substance abuse, self-imposed sleep deprivation, and stress. Training should also include sessions that help management recognize what the company can do to help prevent and eliminate conditions that lead to substance abuse, fatigue, and stress. Other information sessions dealing with lifestyle issues, sleep and sleep disorders, dealing with family and job stress, etc. should be part of the testing program.

5.5 Impact of Fatigue on Performance

A number of studies of the impact of fatigue on transportation personnel have shown that performance degrades significantly when sleep is restricted, and irregular shifts are worked (for an extensive review for commercial drivers see Wylie et al., 1996; for air traffic controllers see Rhodes et al., 1996; for pilots see Rosekind et al., 1994; for mariners see Colquhoun, 1996; and for railroad crews see Circadian Technologies, 1996).

Freund et al. (1995) reviewed policy and past research on fatigue and “loss of alertness”. The paper provides a summary of initiatives undertaken by several federal transportation administrations and agencies. A key finding of the review is the very strong link between accident rates and operator “loss of alertness” caused by fatigue, alcohol, and drugs. In response, the Federal Highway Administration (FHWA) has developed a holistic approach to reducing fatigue in transportation (see Knipling, 1999; and Freund et al., 1995).

Various studies on psychophysiological indicators of fatigue (e.g. Dinges et al., 1998; Rowland et al., 1997; Russo et al., 1999; Trucking Research Institute, 1999; Wierwille et al., 1994; Wylie et al., 1996) show that measures such as percentage eye closure (PERCLOS), eye movement, and pupil dilation are reliable indicators of fatigue. These studies clearly support the use of such indicators as good predictors of the level of alertness (LOA) expected. Performance-based measures can also predict LOA, as shown by Allen et al. (1981; 1990), Miller et al. (1994), and Wylie et al. (1996).

5.6 FFD/RTP Suitability to Transportation Environments

Transportation includes several job classifications with varying degrees of cognitive and psychomotor requirements. The ideal FFD/RTP testing system for one type of transportation job may not be appropriate for others. In fact, the search for the ideal general system to be used for all job classifications may be a “holy grail”. The likely scenario is the implementation of one type of system for jobs that consist of less complex cognitive tasks and a greater requirement for psychomotor performance such as commercial driving, and another type of system that is more suitable for complex cognitive tasks such as planning, troubleshooting, high-level decision-making, and team co-ordination, as found on the flight deck of modern passenger aircraft.

Selection of two or more systems may be necessary, or the customization of an existing system may be the alternative. It may be possible to consider a flexible system that could be modified to suit the specific job classification(s) within a particular work environment.

5.7 Inclusion of Mood, Stress, and Personality Factors

Mood affects performance in that depressed mood can lower motivation (Morris, 1989). However, there is little evidence that mood alone can cause suicidal compulsions, except under conditions of severe depression and/or psychosis (Parkinson et al., 1996).

Stress is considered a strong determinant of aviator and air traffic controller performance, contributing to increased error rates as certain kinds of stress increase (Stokes & Kite, 1994). Increased stress due to increased workload, for example has been shown to correlate highly with error rates (Stokes & Kite, 1994).

Personality factors seem to play a lesser role in affecting overall performance in pilots than cognitive and psychomotor performance (Martinussen, 1996; Martinussen & Torjussen, 1998). In fact, according to these studies, personality factors, academic tests, and intelligence tests show very low correlation with pilot performance. The best predictors of a pilot’s performance were cognitive and psychomotor test performance (Martinussen, 1996; Martinussen & Torjussen, 1998). Notwithstanding, the identification of specific personality flaws may be useful for screening out individuals who may have a propensity for depression or psychosis.

Psychological tests measuring scales for personality traits used to predict flight performance may be confounded by many other factors such as motivation, mood, and present stress (Martinussen, 1996; Martinussen & Torjussen, 1998). On the other hand, better personality tests do exist (Hörmann & Maschke, 1996) and are also computerized, making their implementation easier. Perhaps improved personality tests, combined with mood scales and stress assessments, may help to identify potential risk factors.

5.8 Mode-specific Versus Multi-modal Application

It may be necessary to seek out a system that is suitable for one mode and a different system for another. The use of systems that are specific to one type of job has been described for commercial drivers (Miller, 1996; Stein et al., 1992), for air traffic controllers (O’Donnell, 1992; Thorne, Genser, Sing, & Hegge, 1985), and for pilots (Arnegard, 1990; 1991; Graeber et al., 1990; Horst & Kay, 1988; Schrot & Thomas, 1988; Schlegel & Gilliland, 1992). Further investigation into the application of these systems to marine and rail operations should be considered.

6 Conclusions

6.1 The State of the Art for FFD/RTP Testing Systems

Existing FFD/RTP testing systems show considerable promise as effective screening devices for risk factors such as alcohol, drugs, and fatigue. Many of the systems that are self-contained and consist of cognitive/psychomotor tasks are likely to be more easily applied to the work environment. However, physiological-based systems are validated for sensitivity to some risk factors, but not to all (e.g. stress and mood). On the other hand, some systems that are still in development but are not packaged in a robust, single, stand-alone device, have mood and stress tests that may be adapted to existing systems. Also, the length of time to take the test appears to be a drawback in some of the more sophisticated systems that capture several aspects of performance.

6.2 The Ideal System Versus the Practical System

The ideal system appears to be a combination of a quick physiological measure, a few basic high-level cognitive tests, and some elements that assess stress and mood. Although such a system will satisfy the requirements for complex operations, it would be overkill in some work environments. A practical system would be flexible and configurable to suit the work environment and job classifications used. More than one system may be necessary to meet the demands of the many transportation jobs that have major personal and public safety implications. None of the existing systems comes ready-made with the ability to easily reconfigure, although some systems do allow specific choices of tests and provide for some adjustment for emphasis on types of representative cognitive tasks. Unfortunately, these systems often have longer testing periods, and require detailed analysis of results, offline (i.e. results are not immediate). Adding automated analysis will likely eliminate this problem.

6.3 Implementation of Multiple Systems for the Transportation Environment

Application of the FFD/RTP testing systems within the transportation industry offers challenges to vendors. Some have created imaginative ways to package their systems for mounting in the cab of commercial vehicles, for example. Some provide their systems for both in-terminal and in-cab usage. The latter are specialized systems that are specific to a particular job (e.g. driving a commercial vehicle). On the other hand, if such a system has been fully validated as an acceptable screen for all driving tasks, a large portion of the transportation industry can be served. There may be added value to a multiple system approach because a strong vendor base could be established that encourages competition and continual system improvement.

Considerable work has been done to validate in-vehicle monitoring devices that measure eye lid closure speed, period length, and frequency (see various papers in Caroll, 1999, for reports on this research). Knipling (1999) describes the potential for this technology to provide practical and reliable monitoring of fatigue in the cab during all driving operations. Such systems could be used to augment the in-terminal FFD/RTP testing systems, providing the driver with an advanced warning of impending unsafe levels of fatigue.

6.4 The Holistic Approach – Multiple Systems

Freund et al. (1995) suggest that an LOA program must consider a multi-screen approach, whereby the first screen occurs in the terminal prior to departing, and a second or third screen occurs at other points during the trip. This approach adds the ability to determine the LOA of a driver, pilot, or engineman later in the shift when fatigue may have increased, or when substances may have been taken. Hence, the employee's LOA can be determined and a decision can be made as to whether the employee shall continue or not. Such a dual screening approach would require a portable device that can be installed in the vehicle, and that can maintain a record for review following the completion of the trip, or could be a "less portable" system located at various points along the route.

The in-terminal system can be a desktop system that may also be more comprehensive, providing a more complete initial screen. The in-vehicle system must be more specific and require less test time. This way, the dual screen will encourage the employee to apply fatigue management strategies, abstain from performance degrading substances, and take a break (rest/sleep) where appropriate. Rests and sleep sessions (naps) can be planned into the duty schedule and where possible facilities to accommodate these requirements can be provided. In fact, in some operations, such as airline operations, each terminal could have screening systems installed in pilot lounge.

6.5 An Immediate Requirement for Risk Factor and Job Performance Cross-Validation Studies

Double-blind validation studies must be conducted with a focus on both the risk factors and on job performance. The testing systems must be sensitive to all of the risk factors that pose the greatest threat to personal and public safety, and must translate effectively into actual job performance. It may not be enough to simply screen for the risk factors alone. The ability to perform key safety-related job tasks may not be represented by the test components, and hence job performance may be unacceptably poor even after passing the test. This may be due to the inability of the test to detect synergetic effects between risk factors. The system may lack a test component sensitive to other important risk factors (for example, mental stress), or it may not be able to account for novel high-stress, high-workload situations.

6.6 Optimal Systems for Transport Applications

The following systems have been identified for follow-up assessment and evaluation through pilot and validation studies for specific application to certain transportation job environments. These systems are either ready for use in an operational environment, or will be easily modified to be ready for use in the field. All of the systems rated highly in the assessment (see Table 4-1), and have either been proven for use in transportation or have a strong potential for this type of application.

- *COGSCREEN*™
- *NMRI-PAB*
- *Factor 1000*
- *FIT 2000*
- *ReadyShift* ®
- *ART-90*
- *TOPS*
- *MABT*
- *NovaScan*

7 Recommendations

The following recommendations are derived from the conclusions of this review:

- Profile the transportation jobs where safety is a major concern, so that decisions can be made regarding the application of specific FFD/RTP testing systems.
- Determine the adequacy of appropriate testing systems for each job class and the level of customization that may be required of test components.
- Pilot test systems tailored to specific classes of jobs such as psychomotor driving tasks (e.g. commercial drivers), cognitive team-based operation (e.g. air and marine traffic controllers), cognitive team-based operation combined with psychomotor skills (e.g. air and marine pilots), and one-person operation and alertness monitoring (e.g. railroad engineers, commercial drivers).
- Conduct the pilot tests within selected work environments, using simulation where appropriate.
- Investigate the application of FFD/RTP testing systems to Canadian marine and rail operations.
- Investigate the use of multiple-system approaches involving in-terminal systems combined with portable in-vehicle systems.
- Conduct cross-validation studies on several FFD/RTP testing systems to examine both risk factor and job performance representation (validity) for several transportation work environments.

Table 7-1 outlines systems that should be considered for pilot testing in the listed application areas.

Table 7-1: Systems to Consider for Pilot Testing

FFD/RTP Testing System	Application Area(s)
Readyshift®	Commercial driving
ART-90	Commercial driving
Factor 1000	Commercial driving
TOPS	Commercial driving
FIT 2000	Commercial driving Air traffic controllers Aviation pilots Railway enginemen, control operators, etc. Marine pilots, navigators

Table 7-1 continued

FFD/RTP Testing System	Application Area(s)
Cogscreen	Air traffic controllers Aviation pilots Railway control operators, etc. Marine pilots, navigators
NovaScan	Air traffic controllers Aviation pilots Railway control operators, etc. Marine pilots, navigators
NMRI-PAB	Air traffic controllers Aviation pilots
MABT	Air traffic controllers Aviation pilots

References

Acker, W., and Acker, C. (1982) Bexley Maudsley Automated Psychological Screening and Bexley Maudsley Category Sorting Test: Manual. Windsor, G.B.; NFER-Nelson.

Allen, R.W., Stein, A.C., and Jex, H.R. (1981) Detecting human operator impairment with a psychomotor task. *Proceedings of the 17th Annual Conference on Manual Control* Pasadena; Jet Propulsion Laboratory. Publication 81-95: 611-625.

Allen, R.W., Stein, A.C., and Miller, J.C. (1990) Performance testing as a determinant of fitness-for-duty testing. Warrendale; Society of Automotive Engineers, paper number 901870.

Arneguard, R.J. (1991) Operator strategies under varying conditions of workload. National Aeronautics and Space Administration. Contractor report 4385. NASA Langley Research Center

Attwood, D.A., Nicolich, M.J., and Muise, E.C. (1994) The effects of shift schedules on the performance of control room operators: Implications for a fitness-for-duty test. Presented at the Second International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, Society of Petroleum Engineers, Jakarta, Indonesia.

Bittner, A.C., Jr., Smith, M.G., Kennedy, R.S., Staley, C.F., and Harbeson, M.M. (1985) Automated portable test system (APTS): Overview and prospectus. *Behavioral Research Methods, Instruments, and Computers*, 17: 217-221.

Canadian Centre for Occupational Health and Safety (1987) *Drug Testing in the Workplace: A Conference*. Hamilton, December.

Caroll, R.J. (ed.) (1999) *Occular Measures of Driver Alertness: Technical Conference Proceedings*. Washington, D.C.; FHWA's Office of Motor Carrier and Highway Safety and National Highway Safety Administration

Chiles, W.D., Alluisi, E.A., and Adams, O.S. (1968) Work schedules and performance during confinement, *Human Factors*, 10: 143-196.

Chiles, W.D., Jennings, A.E., and West, G. (1972) Multiple task performance as a predictor of the potential of air traffic controller trainees. Report AM-72-5, Federal Aviation Administration Office of Aviation Medicine.

Circadian Technologies (1996) *Canalert '95: Alertness Assurance in the Canadian Railways*. Cambridge; Circadian Technologies.

Colquhoun, P. (1996) Shiftwork at sea. In Colquhoun, P., Costa, G., Folkard, S., and Knauth, P. (eds.) *Shiftwork: Problems and Solutions*. Bern; Peter Lang.

Combs, R.H. and Jolyan West, L. (1991) *Drug Testing: Issues and Options*. London; Oxford.

Comer, D. (1995) An evaluation of fitness-for-duty testing. Presentation at the 103rd Annual Convention of the American Psychological Association, New York, Aug. 15.

Cook, R.F., Hersch, R.K., and McPherson, T.L. (1999) Drug assessment methods in the workplace. In Mieczkowski, T. (ed.) *Drug Testing Technology: Assessment of Field Applications*. New York; CRC Press.

Dinges, D.F., Mallis, M.M., Maislin, G., and Powell IV, J.W. (1998) Evaluation of techniques for ocular measurement of an index of fatigue and the basis for alertness management. NHTSA report DOT HS 808 762.

Freund, D.M., Knipling, R.R., Landsburg, A.C., Simmons, R.R., and Thomas, G.R. (1995) Holistic approach to operator alertness research. Paper no. 950636. Presented at the Transportation Research Board Annual Meeting. Washington, D.C.

Gilliland, K. and Schlegel, R.E. (1993) Readiness-to-perform testing: A critical analysis of the concept and current practices. Federal Aviation Administration research report DOT/FAA/AM-93/13.

Gilliland, K. and Schlegel, R.E. (1995) Readiness to perform: testing and the worker. *Ergonomics in Design*, January: 14-19.

Gilliland, K. and Schlegel, R.E. (1997) A laboratory model of readiness-to-perform testing: Learning rates and reliability analyses for candidate testing measures. Federal Aviation Administration research report DOT/FAA/AM-97/5.

Glenn, S.W., and Parsons, O.A. (1990) The role of time in neuropsychological performance: investigation and application in an alcoholic population. *The Clinical Neurospsychologist*. 4: 344-354.

Glenn, S.W., and Parsons, O.A. (1991) Impaired efficiency in female alcoholics neuropsychological performance. *Journal of Clinical and Experimental Neuropsychology*. 13: 895-908.

Graeber, R.C., Rosekind, M.R., Connell, L.J., and Dinges, D.F. (1990) Cockpit napping. *ICAO Journal*. Oct.: 6-10

Gustafson, K.H. (1994) *Drug and Alcohol Testing for Local Government Transportation Employees: The Public Employer's Guide*. Washington, D.C.; Shaw, Pittman, Potts & Trowbridge.

Hegge, F.W., Reeves, D.L., Poole, D.P., and Thorne, D.R. (1985) The unified tri-service cognitive performance assessment battery (UTC-PAB) II: hardware/software design and specification. Fort Detrick, MD; U.S. Army Medical Research and Development Command.

Horst, R.L. and Kay, G.G. (1988) Report of comparative studies of cognitive tests: cognitive function evaluation in medical certification of airmen. Report FAA/730-42b, Oklahoma City; Federal Aviation Administration.

Horst, R.L. and Kay, G.G. (1991) Cognitive function evaluationin medical certification of airmen: Development and validation of a prototype test battery. Report FAA/933-014-90, Oklahoma City; Federal Aviation Administration.

Kane, R.L., and Kay, G.G. (1992) Computerized assessment in neuropsychology: a review of tests and test batteries. *Neuropsychology Review*. 3: 1-118.

Kennedy, R.S., Dunlap, W.P., Jones, M.B., Lane, N.E., and Wilkes, R.L. (1985) Portable human assessment battery: stability, reliability, factor structure and correlation with tests of intelligence. Final report NSF/BNS 85001. Washington, D.C.; National Science Foundation.

Kennedy, R.S., Turnage, J.J., Rugotzke, G.G., and Dunlap, W.P. (1993) Effects of graded dosages of alcohol on nine computerized repeated-measures tests. *Ergonomics*, 36 (10): 1195-1222.

Kennedy, R.S., Turnage, J.J., Rugotzke, G.G., and Dunlap, W.P. (1994) Indexing cognitive tests to alcohol dosage and comparison to standardized field sobriety tests. Journal of Studies on Alcohol, 55: 615-628.

Kennedy, R.S., Lanham, D.S., and Turnage, J.J. (1995) Readiness for Duty: Tuning false positives by simulation from empirical data. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting.* 809-813.

Kennedy, R.S. and Drexler, J.M. (1996) Application of multiple cut-offs for fitness-for-duty testing. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.* 569-573.

Knipling, R.R. (1999) Prospectus: The behavioral power of “Alertomometer” feedback. In R.J. Caroll (ed.) *Occular Measures of Driver Alertness: Technical Conference Proceedings*. Washington D.C.; FHWA’s Office of Motor Carrier and Highway Safety and National Highway Safety Administration: 151-159.

Martinussen, M. (1996) Psychological measures as predictors of pilot performance: a meta-analysis. *International Journal of Aviation Psychology*. 6 (1):1-20.

Martinussen, M., and Torjussen, T. (1998) Pilot selection in the Norwegian Air Force: a validation and meta-analysis of the test battery. *International Journal of Aviation Psychology*. 8 (1):33-45.

Merkle, P.J., Kennedy, R.S., Smith, M.G. and Johnson, J.H. (1985) Microcomputer-based field testing for human performance assessment. Paper presented at the Twenty-seventh Annual Meeting of the Military Testing Association, San Diego, California.

Mieczkowski, T. (ed.) (1999) *Drug Testing Technology: Assessment of Field Applications*. New York; CRC Press.

Miller, J. (1996) Fit for duty. *Ergonomics in Design*, 4 (2), 11-17.

Miller, J.C., Kim, H.T., and Parseghian, Z. (1994) Feasibility of carrier-based fitness-for-duty testing of commercial drivers. El Cajon; Evaluation Systems Inc.

Morris, W.N. (1989) *Mood: A Frame of Mind*. New York; Springer-Verlag.

National Transportation Safety Board (1999) *Evaluation of U.S. Department of Transportation Efforts in the 1990's to Address Operator Fatigue*. NTSB/SR-99/01. Washington, D.C.; NTSB.

NTI, Inc. (1995) The effect of alcohol and fatigue on an FAA readiness-to-perform test. ADA299076. Federal Aviation Administration research report DOT/FAA/AM-95/24.

O'Donnell, R.D. (1992) The NovaScan test paradigm: Theoretical basis and validation. Dayton, Ohio; Nova Technology, Inc.

Parkinson, B., Totterdell, P., Briner, R.B., and Reynolds, S. (1996) *Changing Moods: The Psychology of Mood & Mood Regulation*. London; Longman.

Potter, B.A. and Orfali, J.S. (1990) *Drug Testing at Work: A Guide for Employers and Employees*. Berkley; Ronin Publishing.

Powell, D.H., Catlin, R., Funkenstein, H.H., Kaplan, E.F., Ware, J.H., Weintraub, S., and Whitla, D. (1990) The assessment of cognitive skills: a computerized neuropsychological screening battery. Cambridge; Risk Management Foundation of the Harvard Medical Institutions.

Rhodes, W., Heslegrave, R., Ujimoto, K.V., Hahn, K., Zanon, S., Marino, A., Côté, Szlapetis, I., and Pearl, S. (1996) *Impact of Shiftwork on Air Traffic Controllers, Phase II: Analysis of Shift Schedule Effects on Sleep, Physiology and Social Activities*. TP 12816E, Transportation Development Centre report.

Rosekind, M.R., Gander, P.H., Miller, D.L., Gregory, K.B., Smith, R.M., Weldon, K.J., Co, E.L, McNally, K.L., and Labacqz, J.V. (1994) Fatigue in operational settings: examples from the aviation environment. *Human Factors* 36 (2): 327-338.

Ross, L.E. and Mundt, J.C. (1996) Methodological issues in research on the effects of alcohol on pilot performance. *International Journal of Aviation Psychology,* 6 (1): 95-106.

Rowland, L., Thomas, M., Thorne, D., Sing, H., Davis, H., Redmond, D., Belenky, G., Peters, R., Kloeppel, E., and Krichmar, J. (1997) Oculomotor changes during 64 hours of sleep deprivation. *Sleep Research* 1997; 26: 626.

Russo, M., Thomas, M., Sing, H., Thorne, D., Balkin, T., Westenten, N., Redmond, D., Welsh, A., Rowland, L, Johnson, D., Cephus, R., Hall, S., Krichmar, J., and Belenky, G. (1999) Sleep deprivation changes correlate with simulated motor vehicle crashes. *Proceedings of the Ocular Measures of Driver Alertness Conference in Herndon, Virginia,* April 26-27. 119-124.

Schlegel, R.E. and Gilliland, K. (1990) Evaluation of the criterion task set – part1: CTS performance and SWAT data – baseline conditions. Report number AAMRL-TR-90-007. Wright-Patterson Air Force Base, Ohio; Armstrong Aerospace Medical Research Laboratory.

Schlegel, R.E. and Gilliland, K. (1992) Development of the UTC-PAB normative database. Technical report AL-TR-92-0145, Wright-Patterson AFB, Ohio; Armstrong Laboratory.

Schrot, J., and Thomas, J. (1988) Performance assessment battery software. Technical report NMRI-88-6. Bethesda; Navy Medical Research Institute.

Stein, A.C., Parseghian, Z., Allen, R.W., and Miller, J.C. (1992) High risk driver project: theory, development and validation of the Truck Operator Proficiency Systems (TOPS). Hawthorne; Systems Technology Inc. Phase 3 report, Arizona Department of Public Safety.

Stokes, A. and Kite, K. (1994) *Flight Stress: Stress, Fatigue, and Performance in Aviation.* Hants; Avebury Aviation.

Thorne, D., Genser, S., Sing, H., and Hegg, F. (1985) The Walter Reed Performance Assessment Battery. *Neurobehavioral Toxicology and Teratology*, 7: 415-418.

Trucking Research Association (1999) Eye-activity measures of fatigue and napping as a fatigue countermeasure: Final report. FHWA report FHWA-MC-99-028.

Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J., and Kirn, C.L. (1994) Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. NHTSA report DOT HS 808 247.

Wylie, C.D., Schultz, T., Miller, J.C, Mittler, M.M., and Mackie, R.R. (1996*) Driver Fatigue and Alertness Study*. U.S. Department of Transportation, Federal Highway Administration and Transport Canada – Transportation Development Centre report TP1287E.