Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-93/13		

4. Title and Subtitle	5. Report Date
READINESS TO PERFORM TESTING: A Critical Analysis of the Concept and Current Practices	August 1993
	6. Performing Organization Code
7. Author(s) Kirby Gilliland, Ph.D. and Robert E. Schlegel, Ph.D.	8. Performing Organization Report No.
9. Performing Organization Name and Address University of Oklahoma Department of Psychology 1000 Asp, Room 314 Norman, Oklahoma 73019	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. Contract DTFA-02-92P23499
12. Sponsoring Agency name and Address FAA Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	13. Type of Report and Period Covered
	14. Sponsoring Agency Code

15. Supplemental Notes

This work was performed under task AM-B-93-HRR-193 and Contract 92P23499.

16. Abstract

Readiness to Perform (RTP) testing has become an increasingly popular alternative to biochemical screening as a method for assessing risk factors (i.e., drug, alcohol, fatigue, etc.) in the workplace. The focus of RTP testing is on the assessment of worker performance capability prior to job engagement. Thus, RTP testing is a method of assessing the state of preparedness of workers for performing their jobs. Unfortunately, RTP testing is being distributed and applied in the workplace despite a lack of knowledge regarding its conceptual basis and use. This report summarizes a project that investigated the conceptual and methodological foundations of RTP testing. Based on their expertise and background in task battery research, development, and validation, the authors explored the principles and empirical research on which the concept and methods of RTP testing appear to be based. This evaluation included literature reviews, methodological assessments, and an analysis of popular RTP measures. This report first provides a brief review of some of the background for RTP testing and attempts the difficult task of defining RTP. The report then focuses on specific issues and problems in conceptualizing and implementing RTP testing. Among the numerous issues addressed are the more important comprehensive factors of determining the appropriate criteria for RTP prediction and establishing the validity of RTP through various forms of empirical research. In addition, a number of the problems associated with implementing RTP testing are presented. Some examples of these problems are: testing methodology, comparative standards, time limits in testing, general versus restricted populations, and the hidden costs of RTP testing. Finally, this report includes two supporting literature reviews of areas that form the rational foundation for RTP testing. One review provides a summary of some of the major computerized performance assessment batteries from which many of the current RTP measures were derived. The other review surveys research literature on the influence of selected risk factors on human performance.

17. Key Words	18. Distribution Statement
Fitness-for-duty Readiness-to-perform Drug-Alcohol Testing Test Theory Criterion Validity	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	78	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

Readiness to Perform Testing:
A Critical Analysis of the Concept and Current Practices

INTRODUCTION

A growing problem in modern work environments is the presence of workers who are under the influence of alcohol or drugs. Recent surveys and reports have provided ominous insights into what may be occurring in the workplace. Wrich (1988) reported that as many as 65% of individuals between the ages of 18 and 25 years had experienced illicit drugs. Backer (1987) suggested that nearly one in five Americans between the ages of 20 and 40 years had used an illicit drug within one month of the survey. Equally troublesome was a study revealing the involvement of alcohol in nationwide transportation systems (Bureau of National Affairs, 1986). For example, about 30% of railroad employees admitted drinking alcohol on the job in the past year, and 48 railroad accidents in the past decade were believed to be alcohol related. Such findings suggest that the working age population in America is certainly exposed to alcohol and illicit drug use. Exposure occurring in the work environment also seems clear, either through direct use or interaction with those who are intoxicated. The U.S. Chamber of Commerce estimated that drug abuse costs employers in the United States nearly $60 billion a year (as cited in Stone and Kotch, 1989).

In response to the problem of business-related alcohol and illicit drug use, many organizations have implemented drug testing programs. It has been estimated that 50% of medium and large businesses test current or prospective employees for drug use (Guthrie and Olian, 1989). Of those businesses not currently performing drug screening, 10-15% are considering programs in the near future (Bureau of Statistics, 1989). Most of these testing programs utilize some type of biochemical assay, commonly a urinalysis.

While these testing programs appear to provide a useful means of monitoring and discouraging drug and alcohol use in the work environment, they are not without problems. Depending on the type of analysis performed, the reported cost of urinalysis testing ranges from $10 for simple, one-drug tests to several hundred dollars for broad-based screening tests, with the average cost ranging from $25 to $70 (Hanson, 1990; Maltby, 1990). Thus, the expense of drug testing alone is burdensome. And, this type of testing often requires visual observation of the sample collection to eliminate employee deception, thereby adding to the testing costs and employee embarrassment. In addition, biochemical drug screening has not been universally accepted from a legal perspective. The courts have generally upheld the legality of drug screening in occupations that, if compromised by drug involvement, could pose a hazard to the public (Greenfield, 1989; Greenfield, Karren, and Giacobbi, 1989; Sanders, 1989; Sitomer, 1989). However, the courts have not been as uniformly supportive of drug screening for occupations in which public safety is not a central concern. For this and other reasons, drug screening programs provide the potential for significant litigation and its associated costs.

These biochemical assays suffer from other problems as well. Because the tests are selective, screening for alcohol alone will miss individuals who are using illicit drugs, and vice versa. (Broad-based screening involves dramatically increased cost, as noted above.) Biochemical assays may also suffer from inaccuracy -- a number of common prescription and nonprescription drugs mimic the presence of illicit drugs. In most cases, a second stage analysis with a gas chromatograph can be performed to improve the specificity and reliability. These tests also fail to identify *when* the drug was consumed. Because these tests typically identify drug metabolites (and not the drug itself), and because some drug metabolites do not clear the system as rapidly as others, residual traces may be confused with current drug use. In addition, there is a lag, of sometimes up to several days, between sample collection and the availability of test results, a time period that often precludes immediate intervention.

Employee reactions form another source of problems for biochemical assays. Many individuals who are drug tested report feeling that their privacy was violated or feel suddenly mistrusted by their employer (e.g., Hanson, 1990). This may relate to the fact that workers generally believe that medically-related information (such as a laboratory test) is in the private domain (Stone and Vine, 1989). Certainly, the use of direct visual observation in obtaining urinalysis samples provides conditions that could easily lead to a sense of "personal violation." Many employees also fear retribution after a positive drug screen, even if the test was later proven inaccurate (Greenfield et al., 1989; Karren, 1989; Seeber and Lehman, 1989). And, there is some concern about "due cause" issues in drug screening. Drug screening may have the appearance of a "dragnet" approach, especially the implementation of random drug screening methods (see Hartstein, 1987). It has been suggested that drug testing, in the absence of any compelling reason or explanation, appears to have the potential for creating considerable resentment and other negative feelings among employees (see Murphy, Thornton, and Prue, 1991). In fact, the factors cited above may contribute to the finding that drug screening programs sometimes result in decreased worker productivity (Crouch, Webb, Buller, and Rollins, 1989).

One additional problem associated with biochemical drug testing is what this testing method misses. The "risk factors" for job performance do not end with drugs and alcohol. While biochemical testing has the potential for being very effective in detecting drug or alcohol use, it does not assess a large number of other factors that could easily affect work performance. Fatigue, stress, emotional upset or instability, over-the-counter medications, exotic illicit drugs, and common illnesses are just a few of the risk factors that would not be identified in a common drug screen. Yet, these factors have considerable potential for causing significant negative effects on work performance.

In an attempt to protect worker productivity and safety, and to address many of the problems associated with biochemical testing, new approaches to employee drug testing have emerged. Many of the alternative approaches involve performance-based testing techniques. Because these techniques do not have the capability to identify the presence of any specific risk factor, they concentrate on the employee's general level of work preparedness. As a group, these techniques are referred to as "Readiness to Perform" [1] testing methods.

1. Defining Readiness to Perform

Definition: *The term "Readiness to Perform" (RTP) refers to that state in which a person is prepared and capable of performing a job for which the person is willingly disposed and is free of any transient risk factors, such as drugs, alcohol, fatigue, or illness, that might influence job performance.*

This definition assumes some critical prerequisites that form a foundation for capable job performance. First, it assumes that the person has been prepared for the job, that is, the person has the requisite education and training to feel secure in knowing the job requirements. Second, it assumes that, at a more general and

---

[1] Readiness to Perform has also been referred to as "Fitness for Duty," more often in a military context. The term "Fitness for Work" (see Fraser, 1992) has also been used to refer to pre-job physical examinations. The term Readiness to Perform will be adopted in this paper because it addresses a wider range of activities and job-related functions and it does not bear the specific connotations associated with terms, such as *duty* and *fitness*.

enduring level, the employee is physically, mentally, and emotionally suited to the job demands. Third, this definition of RTP assumes capability. It assumes that a person's skills and abilities have been reasonably matched to the job requirements. And fourth, defining RTP includes the assumption that the person is willfully disposed to perform the job. In other words, the person is generally willing and motivated to perform the assigned tasks. Failing to meet any of these assumed factors at least minimally would compromise the capability of performing one's job. Failing to have requisite job knowledge, lacking minimal physical, mental, or emotional capabilities, lacking necessary skills or abilities, or being chronically unwilling or unmotivated to perform a job might all compromise acceptable job performance. These are the factors that form the more enduring foundation of job preparedness. Typically, these enduring factors are assessed and managed during initial job screening, placement, and job training programs. These factors, while playing an important role in overall job performance are not the focus of RTP testing.

Readiness to Perform (RTP) focuses more specifically on those transient risk factors that might lead to a ***state*** *incompatible with acceptable job performance.* Examples of the risk factors that contribute to a more transitory state of job preparedness are alcohol, drugs, illness, and transient motivational factors. Readiness to Perform testing concentrates on detecting the changes in performance that are associated with these risk factors. For this reason, RTP testing focuses on the state of physical, mental, emotional, and motivational preparedness immediately prior to work involvement -- i.e., those personal characteristics believed to be most affected by risk factors, especially alcohol and drugs. In this manner, RTP testing is considered an alternative (or adjunct) to biochemical drug screening. Thus, RTP testing assesses one's performance capabilities prior to actual job engagement with the intent of identifying those individuals who, probably as a result of risk factors, are not prepared to perform their jobs.

2. The Advantages of RTP Testing

According to the vendors of RTP tests, there are decided advantages of RTP testing compared with biochemical drug screening. Many vendors have cited the reduction in cost that RTP testing provides. Because RTP testing usually utilizes fairly simple and rapidly-administered behavioral tests, the cost of administration is believed to be lower. (However, see section on Hidden Costs later in this report.) Another purported advantage of RTP testing is that no specific risk factor is identified. The employee is faced with simply "not being prepared for work," rather than being presented with evidence of specific drug or alcohol involvement. This appeals to workers and trade unions because it reduces invasion of privacy. Some organizations, such as the American Civil Liberties Union, purportedly support RTP testing for this reason. Also adding to the reduction in privacy invasion is the fact that RTP testing does not have the degree of humiliation, embarrassment, or degradation commonly associated with urinalysis collection. The regularity of RTP testing is also more acceptable, thereby reducing the suspicion and apprehension associated with random biochemical drug screening. The video arcade-like appearance of many RTP measures also adds to employee acceptability. Another advantage of RTP testing is that the results are immediate. Employees and management know quickly, and prior to job engagement, whether an employee is prepared for work.

Because RTP testing concentrates on performance preparedness, and not on specifically targeted drugs, it has the potential for reflecting the influence of a much broader range of risk factors. Illness, emotional upset, fatigue, exotic illicit and prescription drugs, and stress, in addition to common illicit drugs and alcohol, can all affect job performance. Reports in the popular press and by at least one manufacturer suggest that RTP testing has been effective in screening for these factors as well (Hamilton, 1991; Maltby, 1990).

Finally, the face validity for job performance of a screening test appears to figure prominently in the level of employee acceptability (Lumsden, 1967; Thorson and Thomas, 1968). Workers seem to accept a screening test more readily if they believe that the test is related to their ability to perform their job. Because RTP measures are behaviorally oriented, they provide what often appears to be greater face validity for job performance. Thus, it would appear that RTP testing has much to recommend it.

3. The Disadvantages of RTP Testing

There are, however, a number of disadvantages to RTP testing. Many of the disadvantages are merely the advantages turned inside out. For example, there is some question about whether a very brief and often narrowly defined behavioral sample is sufficient to assess total job preparedness: Are RTP measures really valid measures of the state of job preparedness?

RTP testing requires repeated behavioral testing. Time spent away from the job includes time for the actual test plus travel time between the assigned duty area and the RTP testing station. Because optimal testing schedules for particular applications have not been identified, there is no clear determination of how much time may be lost from work. In some safety-critical applications, daily testing may be required. Additional concerns include the logistics of test administration, space requirements, and equipment purchases.

Because RTP testing does not provide specific evidence of risk factor involvement, the employer using only RTP testing is left without "hard evidence" of alcohol or illicit drug use in the case of an employee with repeated RTP testing failure. In some cases, RTP tests have been constructed to emphasize the influences of specific risk factors, such as alcohol. But even in these cases, a positive finding would not necessarily confirm the presence of the targeted risk factor. The vendors of RTP tests are well aware of this limitation. However, in spite of cautionary statements made by RTP test vendors, employees often confuse RTP testing with simple drug screening. For that reason, it is conceivable that failing an RTP test could be just as stigmatizing as failing a biochemical test.

This is only a brief discussion of a few of the possible criticisms of the RTP concept. As can be seen, RTP testing provides unique advantages that to be effective and acceptable must be matched to specific testing needs. As with any effective assessment program, RTP tests must match the unique needs and perspective of the consumer. A number of additional issues related to these problems will be raised in the next section of this report.

ISSUES AND PROBLEMS OF RTP TESTING

This section of the report critiques the RTP concept and testing procedures. Special attention is directed toward a critical analysis of the problems and issues that surround RTP. With any new application of existing technology there are always problems and issues that must be resolved. Admittedly, the implementation and validation of any new technique is always more difficult than simple critical appraisal. However, there is a fundamental and proper role for such an analysis -- a type of scientific "checks and balances." This section of the report raises numerous questions, not with the intent of criticizing any specific RTP measure, but rather to aid in the process of stimulating interest and expanding knowledge of RTP concepts and measurement.

This section is organized by topic area. Each topic area addresses a specific RTP issue or problem. The reader should be aware that one charge to the authors was to apply their backgrounds in various areas of experimental psychology, human performance, workload assessment, and industrial engineering to enumerate as many issues and problems as possible related to RTP. Therefore, this list of issues and problems is offered as comprehensive, but perhaps not exhaustive. The reader should also be aware that the authors were asked to provide their collective professional judgments and opinions in evaluating various aspects of the RTP concept. In most cases, the authors have tried to present these judgments and opinions in the recommendations that follow each subsection. These recommendations were prominently placed in boxes to emphasize that they can stand apart from the general critique of the RTP concept and that they do contain the opinions and advice of the authors.

All issues raised here may not apply to all RTP measures. Likewise, not all issues and problems raised here will be of equal merit. The applicability and value of this analysis is derived from applying each point raised to a specific RTP application in question. Therefore, the various issues and problems raised below cannot be viewed as being presented in order of importance. They are, however, ordered to some degree, according to their inclusiveness. Those issues or problems of a more general or pervasive nature are listed first followed by more detailed points.

1. Defining the Concept

Computer-based Readiness to Perform testing is a relatively new concept. While based on decades of human performance research, RTP testing presents a new application of this technology arising from the need to address drug screening more adequately. With this new application goes the responsibility to define carefully the concept of RTP, and the specific techniques used to measure it. Yet, this has not happened. Perhaps it is due to the nascent stage in the development of RTP, or to the variation in terms used to describe this concept, that one finds no clear definition for it in the literature. Nonetheless, a definition of RTP is important because the manner in which RTP is operationalized in the form of an actual test is based largely on that definition. For example, if RTP is defined primarily in terms of physical performance, then the operational RTP measure of choice will probably be more physiological or psychophysiological in nature. If RTP is defined more in terms of effects on mental function, then cognitive measures are likely to be emphasized.

A number of vendors of RTP measures do have product literature available. Among those documents sampled for this report, none clearly defined a concept synonymous with RTP and differentiated it from other more enduring factors related to job performance. Thus, it appears that RTP is a consensually agreed upon area of investigation and application, but it continues to go unclearly defined. It is hoped that the definition provided in this report will serve to stimulate further discussion and refinement. Surely, without some consistency in terminology and definition, the advancement of our knowledge of RTP will be impeded.

> Recommendation. In assessing any proposed RTP testing program, special consideration should be given to the manner in which RTP is defined. If RTP is not clearly defined, then questions should be raised about the linkage between the conceptualization of RTP that is used and the actual RTP measure that is proposed.

2. Needed: A Theory of RTP

General knowledge of the nature of RTP and its measurement needs to be established at the theoretical level. In other words, in addition to having very little in the way of a definition of RTP, there exists even less in terms of a theory of RTP.

The need for understanding RTP at the theoretical level is more than a customary academic appeal. A theory is needed to understand more completely the basic principles of RTP that are operable across numerous work environments. Otherwise, we are condemned to solving each RTP application in isolation, without the benefit of a wider sphere of knowledge of the mechanisms underlying RTP. If pursued in a piecemeal manner, the full range of RTP and its measurement will never be fully understood or applied. Likewise, a more complete understanding of RTP at a theoretical level will provide more effective analyses of specific RTP measures.

Recommendation. In assessing any RTP testing program, special consideration should be given to exploring its theoretical foundation. Have the vendors developed an RTP measure on a firm theoretical base or is it an application not well grounded in theory? At a minimum, the vendors should be able to articulate their conceptualization of RTP in theoretical terms, as opposed to simple, applied terms. They should be able to offer their views on the nature of RTP and where RTP falls in the dynamics of the worker-performance relationship. One should also ask how closely the RTP test is related to the research literature, as discussed in the sections below.

3. RTP and Prediction: What Is the Criterion?

It seems that from the very beginning, *an important issue is defining what one wants to accomplish through RTP testing.* A careful reading of behaviorally-based RTP product literature reveals many responsible qualifying statements to the effect that RTP is not a drug test, it is not an alcohol test, nor is it a test for other specific stressors: fatigue, illness, and the like. What then is it? Most vendors refer to it in terms of job-related impairment testing or performance decrement screening. In this manner, RTP seems to be somehow associated with one's performance on the job. In fact, RTP test vendors often make the claim that their behavioral measures tap the resources common to many job skills, further implying that RTP measures are related to (or can predict) job performance.

On the other hand, what occurs very quickly is the recasting of these behavioral tests as screens for drug and alcohol abuse. The transition from job-related impairment or performance-decrement testing to drug screening is rapid and may appear logical. The logic goes something like this. Typically, vendors cite some form of research that links the effects of drugs or alcohol to decrements on their tests. Therefore, if these tests show the effects of drugs or alcohol, then monitoring for decrements in the RTP test seems to be a logical way to monitor for drug or alcohol use. Now, at once, we have a measure of job-related performance and a detector of risk factors!

In fact, most people probably enter into RTP testing assuming they are assessing both job performance and the presence of risk factors. And, at some level, they may be. If there is any doubt that such assumptions are being made, that doubt is certainly erased in a perusal of RTP test product literature. The merchandising of these tests is clearly within the context of drug and alcohol screening. The behaviorally-based RTP tests are also promoted for their work sample relevance. Unfortunately, close inspection reveals a perplexing problem.

Let's ask again: What is RTP testing? RTP testing is exactly that -- an assessment of one's state of readiness to perform. It reveals the degree to which one can perform a behavioral task (RTP measure), much in the same manner one has performed it in the past. Perhaps it is because such a logical link has been made between RTP measures and job performance skills that one almost naturally assumes that RTP tests predict job performance. In this same manner, these logical links have been made between RTP measures and risk factor effects. In actuality, neither of these relationships is necessarily true. However, they both could be true. Assuming for the moment that simultaneous prediction of job performance and drug presence is possible, what exactly does one want to predict with an RTP measure? Does one want (or expect) to predict work performance? Or, does one want to predict the presence of risk factors (drugs, fatigue, etc.)?

If the goal of RTP testing is solely to predict the presence of risk factors, then an RTP measure that is sensitive to the influence of risk factors need not predict specific job performance variables at all. That is, if one has a reliable RTP measure and, if one has well-conducted validity studies confirming the sensitivity of that RTP measure to risk factors, then one has the critical elements to predict the presence of risk factors from RTP testing. Predicting job performance with the same RTP measure is not necessarily needed, and in some cases could actually be problematic (see below). In other words, if you are trying to detect risk factors, the RTP measure need only have criterion validity for the influence of risk factors. The intent of such an RTP measure is to establish reasonable doubt about the person's preparedness for work and to provide cause for further evaluation.

On the other hand, it may be important to demonstrate that RTP testing is not only useful for the detection of risk factors, but also for predicting job-specific performance. In this case, the RTP test must have some criterion validity for the job as well as sensitivity to risk factors. Job-related criterion validity must be established through well-controlled experimental studies, not through assumptions based on face validity alone.

> Recommendation. The users of RTP testing should have a very clear idea of how they want to use RTP testing. If it is used for drug and alcohol screening, then selection of an RTP measure should emphasize that capability. If predicting job performance is also necessary, then that criterion should also be applied. Ultimately, the successful selection of an RTP test will depend on identifying the proper criterion variable and having an RTP measure firmly grounded in high-quality predictive validity studies.

4. Criterion Validity and RTP Testing

Criterion validity is a central problem for RTP testing. Criterion validity refers to how well a test predicts the specific construct or behavior it is purported to measure. The degree to which an RTP measure is related to either job performance or a risk factor cannot be assumed -- it must be verified empirically. Further, it should be verified by comparing the specific RTP measure in question with actual job performance measures or with task performance measures while in an experimentally-manipulated risk factor state.

Criterion validity cannot be simply abstracted from prior evidence in the research literature. What is referred to here is the practice of citing basic laboratory research demonstrating the effects of various risk factors on human performance of one type or another as evidence that RTP testing in general (and often some specific RTP measure) is also sensitive to these risk factors. Appendix B, in fact, provides examples of research results for alcohol and other drugs. Although this abstraction may seem logical, in practice it should be used to generate hypotheses or trends, and should not be treated as confirmatory evidence. That various memory tasks have been shown to be sensitive to drug or alcohol consumption in the laboratory does not necessarily mean that a specific RTP measure (even one including a memory component) will be equally sensitive. There are a number of reasons for this conclusion. Not all memory tests are equally sensitive to the risk factor, and many times the ability to control and "tease out" such effects in the laboratory are simply not replicable in an applied RTP testing environment. Perhaps an even more compelling reason is that not all tests, even ones constructed to be similar, are alike.

For example, a recent study investigated the consistencies between similar versions of the same task contained in two different human performance task batteries (Schlegel and Gilliland, 1992). This analysis revealed that, in some cases, versions of tasks differing only in what appeared to be inconsequential formatting features of the visual stimuli resulted in noticeable performance differences. If, by simple modifications in format, one alters the nature of a test (for example, making it more simple), the result could easily be to make the task insensitive compared to laboratory tasks for which risk factor effects were found. In short, no level of abstraction from existing literature will provide the same degree of assurance as carefully conducted validity studies. Unfortunately, these studies are noticeably absent for many of the existing RTP tests.

> Recommendation. Any RTP test should be supported by sound empirical studies assessing the criterion validity of the test. If the test is being promoted as an effective method for screening drugs, alcohol, or any other risk factor, there ought to be clear evidence that the risk factors identified have been shown to influence performance on the RTP test. The scientific credibility of any RTP measure must be very carefully scrutinized. The vendor of an RTP measure should be able to provide completely documented, competently performed investigations that verify the validity and the usefulness of the proposed measure. Preferably, this documentation should rest on research published in archival journals. Minimally, such evidence should be complete enough to be examined for its scientific credibility. There is nothing inappropriate with demonstrating a firm foundation of past research results that supports the general use of any RTP measure. However, any specific RTP measure ought to have criterion validity studies of its own and these ought to be fully documented and readily available for evaluation.

5. Needed: Research on RTP

One outgrowth of this report was the discovery that very little research has been conducted on RTP, and even less has been reported in the open literature. In the course of preparing this document, several computer searches and traditional reviews of scientific and popular literature bases were completed. Few citations for RTP or associated terms were found among the articles searched. However, a number of articles have been published in the popular press on behaviorally-based drug screening. It is possible that little to no research on this concept proper exists.

Or it is possible that none of the research conducted thus far has been published in the open literature. Perhaps both of these explanations are true in the case of RTP.

Certainly, there is a substantial body of literature on the effects of drugs on human performance. But, for a number of important reasons, this research is not the same as well-constructed research studies on RTP measures. It appears that the research that does exist on RTP has primarily been conducted by RTP vendors to support the efficacy of their products. Unfortunately, the claims of such research are too often supported by brief abstracts of these studies in product documentation -- abstracts that do not allow sufficient detail to evaluate scientific merit. Vendors also base claims of RTP efficacy on "proprietary" research that they decline to circulate openly. Understandably guarded within the harsh competitive world of business, such research, while perhaps competently performed, is functionally worthless to the larger research community and to the wary consumer, as well.

> Recommendation. If RTP testing is to be accepted in the long term, more research on the efficacy of specific RTP measures needs to be made available for scientific scrutiny. More basic research needs to be conducted to explore the fundamental principles of RTP and its measurement.

6. Face Validity and RTP Testing

Another area of potential confusion in RTP testing is the issue of face validity and the manner in which it is applied. Traditionally, face validity refers to whether a test appears on the basis of outward appearance to measure what it is purported to measure. Thus, whenever face validity is of concern, it ought to be in reference to the construct being measured by the test in question (see Section 3 above, RTP and Prediction: What is the Criterion?). In looser terms, face validity is sometimes used outside the usual psychometric manner of establishing the linkage between test and criterion to simply describe the overt appearance of a test. In this manner, tests are said to have face validity for a construct if they simply look as if they measure that construct.

Because most RTP measures are implemented to screen for risk factors, the traditional use of face validity ought to refer to the extent that the test appears to measure the influence of risk factors. However, face validity, as applied to RTP testing, is almost invariably in reference to whether or not the RTP test appears to measure job performance. It should be remembered that, for an RTP test to be effective as a drug and alcohol (or risk factor) screen, it need only predict the presence of those factors. It simply does not need to have face validity for job performance to operate effectively in that manner. As an example, a very extensive, prohibitively expensive biochemical test administered every day would have very high predictive ability for the criterion of drug screening, and have very high face validity for drug and alcohol screening -- and have no face validity for job performance. It is quite possible to have an RTP test with the same characteristics. Nor does an RTP test need to have high face validity for risk factors to effectively predict them -- as in the manner of any disguised test. In fact, one danger is possibly reducing the predictive power for risk variables of an effective RTP measure by demanding that it have non-essential job-related face validity.

Very few of the RTP measures on the market provide any data for job-related criterion validity. Many RTP vendors suggest that their tests have some relationship to job performance, but few validate that claim with research. At the same time, most vendors at least suggest a relationship between behavioral RTP measures and job performance. They often support this contention with a "shared-factors" explanation, i.e., both spheres of behavior share skills, resources, abilities, etc.

So why be concerned about job-related face validity? First, there may be some legitimate concern about job-related *criterion* validity, and face validity often accompanies it. The principle advantage is that if an RTP test predicts risk factors *and* job performance, then one may be in a stronger position to defend actions taken to prevent employees from working after a "positive" test result. (More will be said about this in the next section.) However, job-related face validity alone does not increase this potential, nor does it ensure job-related *criterion* validity.

Second, the reference to RTP face validity, *as related to job performance*, often appears to be oriented toward addressing issues other than validity, *per se*. This concern appears to arise from unrelated, yet often quite legitimate, factors such as employee acceptance or other ancillary restraints on testing methods. One main concern with job-oriented face validity and RTP measures appears to be the belief that employees won't accept an RTP measure unless it looks like it measures job performance. There is some evidence to support this view. It has been noted earlier that employees seem to object to tests that do not appear to be related to the abilities necessary for performing their jobs (Lumsden, 1967; Thorson and Thomas, 1968). In addition, anecdotal evidence from aviation research and pilot selection, as well as other areas, suggests that cooperation from subjects is best if there is an obvious link between the test and job-related skills and abilities. The danger here is in not realizing that the face validity of an RTP test for job performance may have nothing to do with its ability to screen for risk factors. If the absence of job-related face validity produces a lack of compliance or support among employees for RTP testing, then perhaps the wrong message was provided the employees in the first place. In general, RTP measures are not designed to test job performance. They test performance preparedness and, by extension, the possible influence of risk factors on that preparedness. From this standpoint, they have excellent face validity. Again, confusion by employers about what is being predicted may lead to false presumptions about face validity.

While face validity for job performance seems to increase the acceptability of the RTP test among workers, it could conceivably be a source of confusion or produce a morale problem if not carefully introduced. For example, workers may assume that the RTP task has predictive validity for job performance based on an apparent high degree of face validity. They may later feel betrayed if they find out that the RTP measure has only face validity for job performance and little or no job-related criterion validity.

Finally, other ancillary forces may place demands on RTP tests for job-related face validity when none is really needed. There may be some reason to require job-related face validity based on legal defensibility; however, in this case, one would prefer clear evidence of criterion validity. The sheer need to overcome management and employee skepticism regarding the test may be a legitimate reason for selecting an RTP test with at least some level of job-related face validity. Also, unrestricted requirements from organizations, such as professional associations or unions, may play a role in the decision process. The important fact to remember is that the existence of job-related face validity does not ensure the ability to actually predict job performance and does not necessarily increase the ability to screen for risk factors.

Recommendation. Define clearly the actual criterion variable for RTP testing in any specific setting. Assess face validity in relation to that criterion variable. Assume that risk factor assessment is the key criterion in most cases, then assess the need for job-related face validity. Consider whether education of employees and management might overcome resistance created by a lack of job-related face validity. Only then, consider altering the task.

7. Risk Factors or Job Performance: What's More Important to Predict?

It has been noted in previous sections above that if risk factor screening is the chief goal of RTP testing, then one ought to select an RTP test with risk factor-related criterion validity. In this case, an additional question is whether the inclusion of job-related criterion validity is also important. This section presents a discussion of some of the relevant issues related to the interrelation of these two sources of criterion validity.

Most RTP testing occurs within the context of seeking a method for risk factor screening. For this reason, the consideration of risk factor-related criterion validity seems self-evident. The problem seems to center on the degree to which job-related criterion validity is also needed. To clarify this problem, let's examine some situations in which the two types of validity do or do not exist. The figure below helps to illustrate some of the potential relationships.

In each case, three elements exist: RTP test, risk factors, and job performance. It is assumed in all cases that risk factors influence job performance in some manner (dark arrow on the right). The influence of risk factors on job performance has been established in some cases through documented evidence, and in other cases it has been assumed. This model also assumes that risk factors influence RTP measures to varying degrees (dark arrow on the left). While risk factors are assumed to influence RTP measures in general, that does not mean all RTP measures are equally effective in predicting the presence of any specific risk factor. Table 1 summarizes the characteristics of three specific cases of interest.

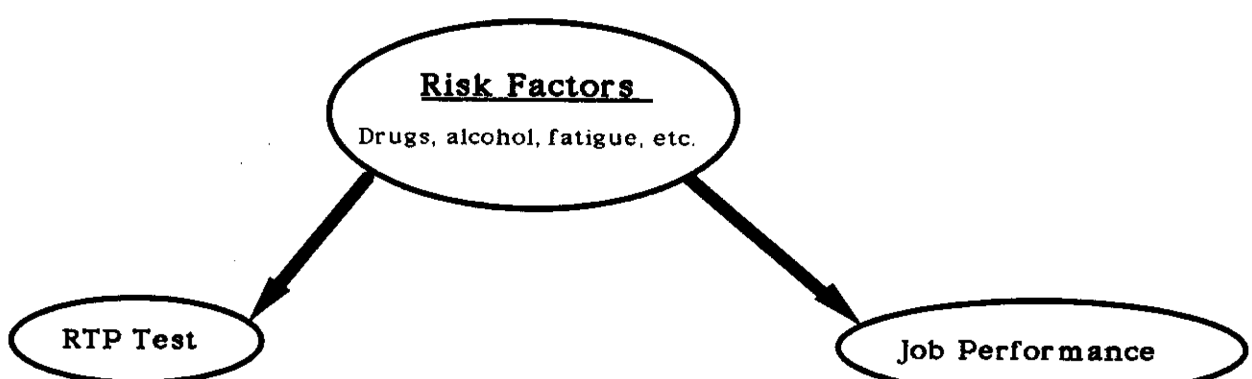
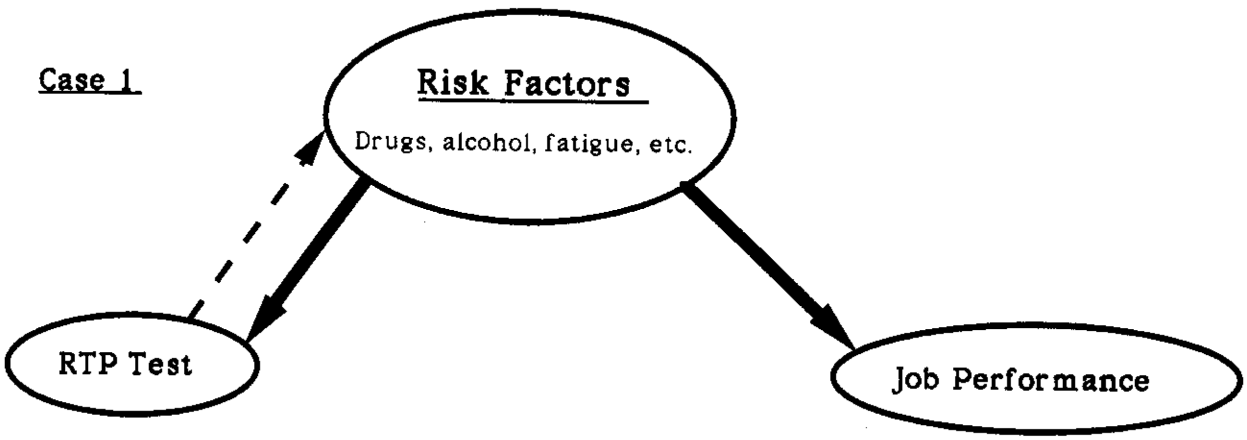


Table 1. **Predictive Validity for Risk Factors vs. Job Performance.**

	Predictive Validity?	
	Risk Factors	**Job Performance**
Case 1	Yes	No
Case 2	No	Yes
Case 3	Yes	Yes

Case 1 below depicts a situation where one has an RTP test and it has criterion validity for (i.e., predicts the presence of) risk factors, as represented by the dashed line. Assume that this RTP test does <u>not</u> have criterion validity for job performance. In this case, the RTP measure can function validly as a screen for risk factors. In other words, one can be assured, with a reasonable degree of confidence (related to the strength of the risk factor-related criterion validity), that significant variation in RTP performance suggests the presence of a risk factor. Obviously, it does not identify the specific risk factor, only that something is preventing the worker from performing in a usual manner.

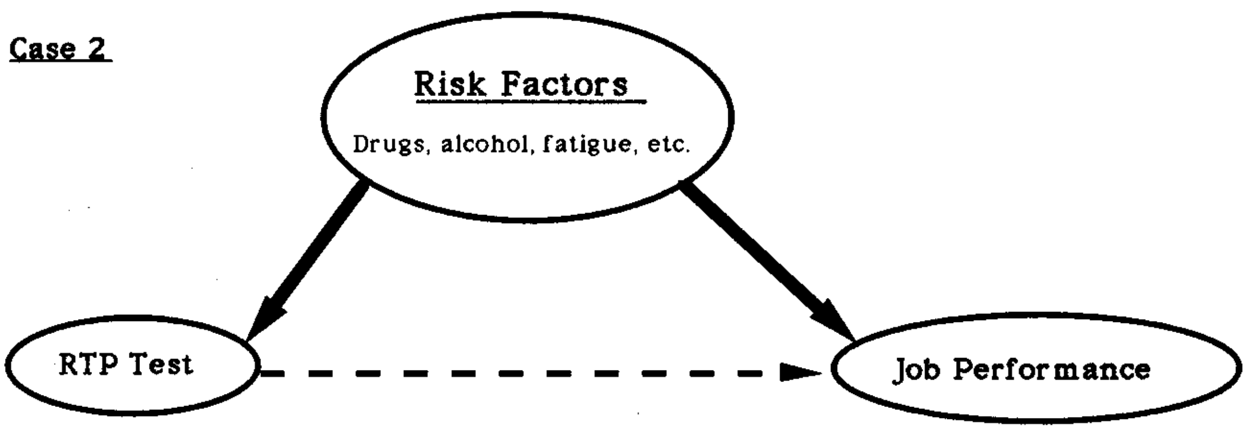


Thus, the presence of degraded RTP test performance in this case suggests the influence of a risk factor. Because risk factors are often assumed to negatively influence job performance, there is an <u>assumption</u> of associated negative job performance capability. The effects on job performance can only be established indirectly in this case. Even though the RTP test is behavioral, without direct evidence of job-related criterion validity, inferences regarding job performance can only be assumed.

This case represents the situation in which many RTP test users find themselves. They believe their RTP measure provides some degree of prediction for risk factors and use that information to protect the integrity of job performance. This type of RTP application is probably best suited to situations where workers vary a great deal in RTP test performance (i.e., there is a wide range of ability in performing the RTP test) and where workers vary a great deal in job performance ability. In such cases, the wide variation in RTP performance will provide better individualized predictive capability for risk variables and avoid problems that may be associated with differences between workers in job performance (see next section). This case also seems well-suited to situations where there is a wide range of job classifications. No single RTP measure can be expected to predict equally well a large number of jobs that may vary considerably in requisite skills and abilities. Maximizing the prediction of risk variables may be much more advantageous.

Case 2 presents a situation where RTP testing has well-established criterion validity for job performance, but no established criterion validity for risk factors. Admittedly, this case might be unusual, given that most RTP testing is predicated on a need to predict risk factors. However, in the case where an RTP test has very little scientifically verifiable evidence of criterion validity, a high degree of job-related criterion validity may provide a valid foundation for its use in risk factor assessment. In this manner, a significant variation in the RTP measure would suggest a more direct inability to perform the job.

Case 2

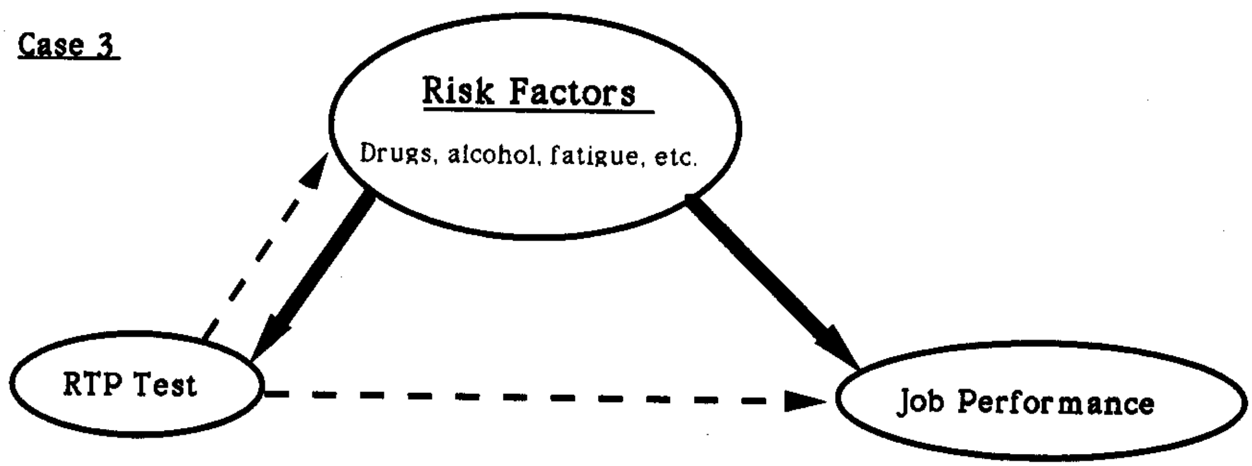


Because we are assuming that job performance, in all cases, is subject to the negative influences of risk factors, such a test result would raise suspicions that some risk factor is affecting performance much like that demanded on the job. In this case, the known or assumed influences of risk factors on job performance are more critical. This type of situation might be well-suited to occupational settings where workers are highly selected for job performance. As a result of such selection, their job performance will probably have less group variability, as will the RTP measure. Significant changes in RTP performance will probably be well outside the general range of group performance and will suggest obvious unpreparedness for work. Even in this situation, there is no substitute for the RTP test having a significant amount of risk factor criterion validity.

Case 3 provides an RTP measure with criterion validity for both risk factors and job performance. In this case, one can be reasonably assured that significant variation in the RTP test suggests unpreparedness due to potential risk factor presence and probable job performance decrements. Due to its increased predictive capability, this case might be used best when decrements in job performance could result in serious property loss or threats to public or personal safety.

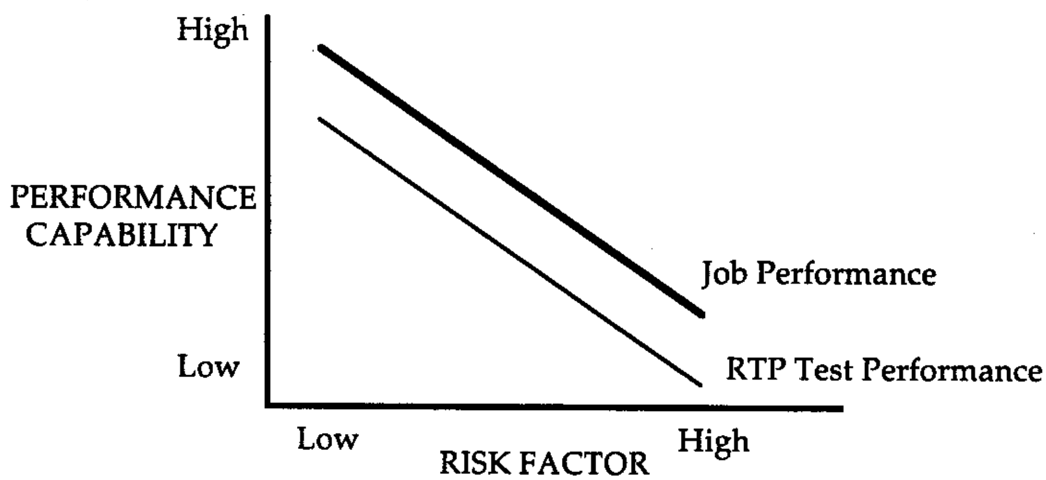
It might be assumed that Case 3 presents the best approach. Again, caution is warranted. Each case presents different advantages and disadvantages. One must approach the method for RTP testing with exactly the same question asked when one selects an RTP test. That is, what is being predicted? In general, Case 3 does present the most potential for predictive power, but only if optimal RTP measures are adopted. Utilizing the Case 3 approach with RTP measures having poor criterion validity would not be as effective as using the Case 1 or Case 2 approach with a highly predictive RTP test. Also, there are some situations where the ability to predict job performance might be a disadvantage (see next section).

Case 3



There is one additional issue that should be considered when evaluating the locus of prediction for RTP tests. As noted above, for RTP tests to be effective, they must have criterion validity for risk factors. One usually assumes that job performance covaries with performance on the RTP test -- both being improved or degraded with the introduction of a specific risk factor. What is critical to understand is that raw RTP test score and relevant job performance indices may be totally unrelated for a group of workers at any given level of the risk factor. However, there may be a very strong relationship between *changes* in RTP test performance and parallel *changes* in job performance.

For example, simple visual reaction time might be very sensitive as an RTP test with respect to the effects of some specific drug. It might also be very poor as a predictor of job performance in a variety of jobs where speed of response is not important. However, as the level of the drug is increased, there may be very pronounced declines in both RTP test performance and job performance. This is simply a situation where the apparent correlation between two variables is being produced by a third underlying variable. While this relationship may be quite complex, it can be represented simply in the figure below.



In summary, the absolute scores on the RTP test and relevant indices of job performance can be totally unrelated at any specific level of the risk factor. However, the manner in which the RTP test changes in response to the risk factor may be very predictive of the manner in which job performance changes in response to the risk factor as well.

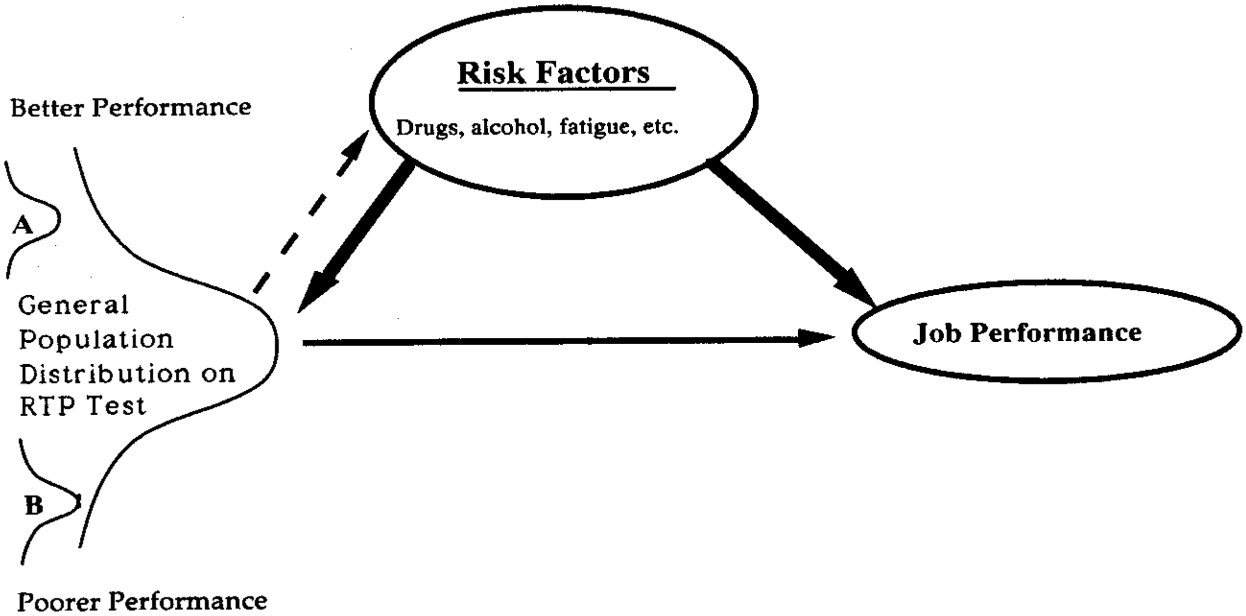
Recommendation. When assessing an RTP test, consider the need for both risk factor-related criterion validity and job-related criterion validity. If only one type of validity is needed, then select an RTP test that optimizes that form of validity. If both types are needed, assess the research evidence for both, given each of the candidate RTP tests. Then, weighing both the need for each type of validity and the evidence for each, make an optimal trade-off decision.

8. Predicting Job Performance from RTP Tests: Individual Differences

In the previous section, it was noted that there are cases where absolute scores on an RTP test might not be related to job performance measures. However, there are cases where raw RTP test scores are substantially related to job performance measures, and these cases may raise special problems. This section addresses one potential problem, the impact of individual differences, and its possible manifestation in relative differences on both RTP test performance and job performance.

The decision to include job-related criterion validity for an RTP test is an important one. On the surface, it might appear that job-related criterion validity would simply add predictive power to the RTP test. In some sense, it does just that (although see section above for conditional statement on optimal measures). However, the addition of job-related criterion validity may not always be desirable. Adding job-related criterion validity to an RTP test increases the direct relationship between the RTP test and indices of job performance. It is conceivable that in certain instances having an RTP test with a strong relationship to job performance may be a disadvantage. In other words, in some cases it may be an advantage to predict risk factors accurately without involving job performance.

One situation where job-related criterion validity might not necessarily be helpful is in cases where employees vary greatly in their RTP and job performance. There is undoubtedly a normal range of acceptable performance for any RTP task. On the figure below, this normal range of RTP test variability is illustrated by the larger normal curve labeled "General Population Distribution on RTP Test." RTP vendors have astutely recognized this fact, controlling for it by using each employee's own rolling average as the basis for comparison. In this manner, employees are never subjected to *a priori* or capriciously developed standards that do not reflect their unique performance capability. However, one major problem still remains. The problem arises because an RTP test with substantial job-related criterion validity now not only has the potential for revealing something about the presence of risk factors, but also reveals something about the manner in which the person can perform the job. If the RTP test has criterion validity for job performance, then it predicts job performance -- it becomes a measure by which workers can be compared with regard to their potential for performing their work. While this may not be advisable, a substantial correlation between the RTP test and well-verified measures of job performance could provide the opportunity for formal comparisons by management or casual comparisons by co-workers.



This issue could become problematic when some people performing a job are a number of standard deviations apart from the performance of co-workers on the same RTP task. Remember that one's RTP test standard is based on one's own prior sample of RTP test performance -- that is, a self-referenced norm, as compared to a group norm. Each person has a distribution of scores, but where those scores fall in relation to everyone else will be different (i.e., reflecting individual differences in RTP test performance). Two such individual distributions of scores are represented by the letters "A" and "B."

As noted in the figure, person A normally performs two standard deviations above the <u>group</u> mean on the RTP task and person B performs two standard deviations below the <u>group</u> mean. This establishes a large absolute difference between these two employees. But, remember also that if the RTP test has criterion related validity for job performance, then this difference also suggests a significant difference in the way each performs the job.

Now consider that on a specific work day, person A comes to work and scores two standard deviations below <u>his/her</u> usual mean performance level. This is based on the <u>self-referenced</u> mean and standard deviation. Of course, a variation of two or even three standard deviations on a self-referenced basis will be a much smaller change in score than if the person changed two standard deviations based on the group mean. Surely, such a performance difference would trigger an alert, suggesting the possibility of risk factor influence. For the sake of the example, person A takes the RTP test again and fails to score in an acceptable range again. Assume also that person B performs as usual, two standard deviations below the general population mean, but stable enough on this day to pass the RTP test. This situation could lead to prohibiting person A from working and allowing another person, who scores much lower in absolute terms on the RTP test, to work. This might sound reasonable on the basis of the "negative " RTP test results. But remember, RTP now reflects job performance as well. Even though person A is performing poorly on the RTP test (with respect to the personal standard), this person is still performing better than person B by a substantial margin on an index of job performance. Such a situation could lead to inequities if a clear relationship between RTP testing and job performance is not defined, or if contingencies are not planned.

Such a situation is difficult to resolve, given the current state of knowledge of RTP testing. It could be that higher absolute RTP test scores, even in the presence of a risk factor, may reflect higher performance on job-related indices -- thereby leading to real inequities. On the other hand, even though a person's absolute RTP test score might be higher than another person's score, it could be argued that degraded RTP test performance for a given individual may reflect degradation in the basic processes underlying decision and judgment skills. Such impairment, regardless of absolute RTP test score, might have catastrophic effects on job performance. It might also be the case that such impairments are manifest primarily during critical events. Thus, on a day when they are impaired, higher absolute RTP test performers might be able to perform as well or better than those scoring lower on the RTP test, *provided there is routine operation of the job*. However, if critical events arise, these workers may be considerably worse in job performance. One might also argue that clear performance variation on the RTP task is still evidence of possible risk factor influence on performance and justifies removing the person from work. However, this logic assumes a position much like that taken with regard to biochemical screening -- namely, that the courts will support such action in the case of safety-sensitive jobs. Unfortunately, no court decisions as of this time have irrefutably supported RTP testing of such employees in the same manner as they have biochemical screening. Until that time, RTP testing remains vulnerable to challenges based on quantifiable and verifiable individual differences in performance.

This situation demonstrates that the greater the predictive validity that the RTP test possesses for job performance, the greater the significance of individual differences on the RTP test. If an RTP measure has substantial criterion validity for risk factor effects, but no real criterion validity for job performance, then how one does in relative terms on the RTP task has no implications for the job. However, if the task also has a high degree of criterion validity for job performance, then day-to-day performance on the RTP task may not only provide information about risk factor presence, but also about how well one can perform the job. *Relative* performance on the RTP task, therefore, becomes meaningful in this situation.

Recommendation. In any RTP testing situation, the value of having job-related criterion validity must be weighed in light of the disadvantages that large individual differences might present. The vendor of any RTP test should be able to document not only criterion validity for risk factor assessment, but also criterion validity for job performance. The consumer must then make an intelligent decision as to what degree they want these validities represented, given their advantages and disadvantages. Note: In many cases, what consumers of RTP tests seem to want is *face validity* for job performance to increase employee cooperation with the RTP testing program. It is possible to have face validity for job performance without having significant criterion validity. In this situation, the RTP test just appears to predict work performance but, in fact, does not have substantial correlations with work indices. This would be one way to solve the problem highlighted in this section. Another solution would be to *select* workers based on job criterion measures. This would have the effect of restricting the range of scores so that all workers would then occupy a much smaller range on the group distribution. Significant deviation from the usual self-referenced RTP standards would be more likely to place the person outside the range of acceptable job performance for many employees in the group.

9. Reliability and RTP Tests

The issue of test-retest reliability and differential stability of RTP tests is rarely raised in the available product literature. This is important in that reliability is directly related to validity. If a test fails to have substantial reliability, then its chances for achieving most forms of validity are poor. Therefore, establishing an acceptable level of reliability is essential for any RTP test.

Recommendation. Be sure to ask for both reliability and validity estimates for any RTP measure being considered. Also, ask for details regarding the studies on which those estimates are based. Are the estimates based on existing literature using a task much like the one being provided? Or, are they studies using the actual RTP test being provided? Do the subject samples sufficiently represent the population of intended use and are they large enough to make reasonable interpretations? Narrow samples (e.g., pilot trainees, power station trainees, or other groups restricted in range) may not provide accurate reliability estimates (underestimated due to restriction in range). Low sample sizes can also result in unreliable correlations, the main statistical test used to establish reliability.

10. Comprehensiveness of the RTP Testing Program

The comprehensiveness of an RTP testing program should be questioned from the very beginning. Is the intent to establish a narrowly evolved program that is directed toward answering a very circumscribed risk factor problem? Or, is this a program based on a more general approach to RTP that will provide, not only a possible answer to a specific problem, but also a broader view of RTP problems and needs within the employment setting? In other words, will this program solve a very narrowly-defined screening problem and have to be duplicated if variations of that problem occur in the future, or will it provide broader insights into larger classes of management problems?

The issue of comprehensiveness can be seen in an analogy to the biochemical screening approach. A urinalysis screen for alcohol will address that one problem, but will miss every other psychoactive chemical agent. What may be more desirable is a screening test that will address more than a single problem and perhaps even provide insight into the dynamics of the problems and remediation methods, as well.

The behavioral approaches to RTP testing seem most promising in this regard, especially if they are linked to more extensive, secondary assessment systems and employee assistance programs (however, see next section below). Even so, if an RTP test is not integrated within a well-constructed theory of RTP, it may fall seriously short of its potential.

Recommendation. Assess not only the theory of RTP testing behind the test being offered, but also the breadth of effective risk factor prediction. If RTP testing is being implemented to deal with a very specific risk factor problem, such as alcohol in the work place, then the RTP test should be maximally predictive of that risk factor. (And, its validity studies should support this.) If broader screening is desired, then the RTP test should have demonstrated capability (validity studies) to predict other risk factors as well.

11. Can a Brief RTP Test Detect Potential Risk Factors?

One assumption that appears to be made is that brief task performance samples such as RTP tests will reveal decrements that constitute evidence of risk factors. There is actually a fair amount of evidence that suggests this may be plausible. There is a considerable amount of research on the influence of stressors such as drugs, heat, sleep loss, etc. on simple task performance. Much of this research suggests that simple performance tasks can be sensitive to the influence of these variables. It is presumably this body of literature that forms much of the foundation for the RTP concept.

However, what is not clear is whether any specific RTP task is sensitive to all or even most of these variables. For example, one task may be sensitive to certain drug effects, but may be relatively insensitive to fatigue or stress. Consideration should be given to the sensitivity of the RTP test, in general, for detecting possible risk factors. This is typically established through validity studies.

Incidentally, this same question can be asked about job performance. Much of the research relating human task performance to job performance is mixed. Some studies (see Cronbach, 1970, or Wiggins, 1973, for examples) have been fairly successful in predicting job performance from simple task performance. These are usually cases where the job task is similar to the screening task. Other studies suggest that brief (e.g., three to five minute) samples of presumably relevant performance tasks predict job performance modestly (e.g., studies on pilot selection; see Blower and Dolgin, 1991). Thus, the fundamental question of whether one can predict more complex job performance from simple tasks is far from answered.

Recommendation. The vendor should be able to provide validity studies verifying those risk factors (or job criteria) for which the RTP test is sensitive. Again, consumers must evaluate these validity studies with respect to scientific credibility and their specific needs..

12. Significant *Improvement* in RTP Test Score

There is at least anecdotal evidence from our lab and one RTP vendor that suggests that some subjects actually **improve** their performance, as compared to previous baseline measurement, under some levels of some risk factors. Again, it should be recalled that these are brief trials. While performance may show improvement in a single, three-minute RTP test trial, that may not be the case with extended job performance. Such improvement under risk factor conditions that would presumably lead to poorer performance is puzzling, yet may be a function of such factors as unique arousal states, unusual focusing of resources, or possibly, performance-enhancing drugs. These experiences suggest that RTP testing may not be a matter of simply detecting decrements in performance. Changes in baseline performance in either direction should be considered as important clues in detecting risk factors.

> Recommendation. While vendors may claim legitimate proprietary rights to RTP test scoring algorithms, they should still be able to provide information regarding the degree or even the manner in which measures of central tendency and/or variability are used in scoring. Certainly they should be able to relate whether variation in one or both directions is considered.

13. Comparability of Risk Factor Influences on RTP Tests and Job Performance

Another question of importance is whether risk factors that are known to cause decrements in laboratory-based human performance tasks cause similar decrements in both job and RTP test performance. Do risk variables (such as drugs, alcohol, fatigue, stress, etc.) that probably affect job performance, affect RTP performance to the same degree and in the same manner? For example, high levels of caffeine consumption (or caffeine withdrawal) can cause significant psychomotor tremor, perhaps enough to negatively affect an RTP test. Yet, jobs requiring more gross psychomotor performance might not be affected. Also, we often hear about people who have consumed alcohol on the job, yet hold and perform their jobs for years without mishap. Often, if it were not for additional environmental stressors or coincidental and unlikely combinations of events, these people might appear totally capable of performing their jobs. Perhaps a better example is provided by people who are emotionally distraught, but can put the problem aside mentally for a few minutes to take the RTP test. Yet, after hours of monotonous work, the problem preoccupies them and they become sufficiently distracted to place themselves and others at risk. The major point is that risk factors may differentially affect job and RTP task performance.

> Recommendation. The available validity studies on any RTP test should provide enough information to determine to what degree the RTP test is consistent with job performance in registering the effects of risk factors. This is, admittedly, a stringent requirement for the relatively new RTP tests. However, this should be an important concern for those who would use these tests.

14. Setting Standards for Acceptable and Unacceptable RTP Performance

Research is needed to determine exactly what constitutes acceptable and unacceptable performance on the RTP measure. Unacceptable RTP test performance is often as simple as a score that varies by an almost arbitrary standard of 1.5 or 2.0 standard deviations from baseline. How does a score variation of 1.5 standard deviations differ from a score variation of 2.0 standard deviations? The vendor should be able to offer an explanation. And, this explanation should be based on something more than just the properties of the normal distribution. For example, the vendor can easily say that a score deviating 1.5 standard deviations from the mean has a certain low-level probability of occurrence based simply on the normal curve distribution properties. But, the important question is not simply the probability of occurrence, but the probability of occurrence in the presence of risk factors. Employers usually want to know their likelihood of detecting an impaired employee, not normal variation. It would seem that a standard based on the individual's standard deviation would place at a greater disadvantage the consistent performer over the erratic performer who has a much higher standard deviation. Greater performance latitude in an absolute sense is allowed in the case of the erratic performer before the person is deemed to have "failed."

> Recommendation. The consumer should be involved in the standard-setting process from the initial establishment of an RTP testing program. The consumer should consider the desired accuracy in predicting the presence of risk factors and weigh that need against the cost of screening. The vendor should be able to provide data to verify the prediction ability of the RTP test at various performance standard levels (e.g., 1.0, 1.5, 2.0 standard deviations from baseline) given at least a few representative risk factors (such as alcohol or sleep loss).

15. Equating RTP Testing with Drug Screening

In spite of common sense and cautionary statements made by RTP test vendors, RTP testing is often viewed as synonymous with drug screening. It is an easy and logical next step, especially for a layperson, to view it as equivalent to biochemical screening. Why does this happen? One reason is probably that RTP tests are promoted as "alternatives to drug screening." The term "alternative" is probably not perceived as something in place of biochemical testing (suggesting a difference), but simply interpreted as a "substitute" (suggesting comparability). If there is doubt that this linkage happens, a cursory reading of any of the popular press articles on behavioral drug screening measures will eliminate the doubt. For example:

> *"Performance testing often detects instances of drug use that fall through the cracks in urinalysis....Many...employers, especially those that employ people for safety sensitive jobs, have relied on drug testing because they know of no alternatives. Performance testing may well be that alternative."* (Maltby, 1990)

> *"Using random drug testing to promote workplace safety is an issue that has been bedeviling employers and civil libertarians. Now, ...* [there is] *a simple, computer-based test that could go a lot further toward determining an employee's fitness for work than drug tests ever have."* (Hamilton, 1991)

The important point to be emphasized is that RTP testing is not drug screening in the strict sense. It is, at best, risk factor screening in the broadest sense. And, more accurately, it is screening for performance preparedness.

> Recommendation. Carefully assess the manner in which vendors present their RTP tests. Is there an unwarranted transfer from performance readiness to drug screening? Is there evidence of logical leaps that can not be substantiated? People are looking to RTP tests as alternatives to drug screening. Is there a palpable realization of the limitations of RTP testing in this application?

16. The Impact of RTP Testing on the Worker

One question that has not been dealt with very clearly in the scarce literature on RTP testing that does exist is the influence of RTP testing on the employee. There are some reports (Maltby, 1990; Murphy et al., 1991) that suggest employees like RTP tests better than biochemical tests. This is probably because of the sense of personal violation related to biochemical testing, as well as its more definitive self-incriminating nature (i.e., positive proof of abused substances). In addition, the video-game nature of most RTP tests appears to appeal to employees, and employees often see behavior tests as having more face validity for their jobs. So, in some ways, RTP tests have some very positive aspects.

However, the consequences of "failing" an RTP test seem to be studiously avoided by RTP test vendors. What is usually presented is a scenario in which the worker is informed about being in or out of performance bounds. If they are outside, they need to take the test again or consult a supervisor. This is all couched within humane and considerate dialogue, assuredly to protect the employee's dignity.

However, there may be some obfuscation here as well. Surely, failing to "pass" the RTP test and not proceeding on to work along with other workers will be noticed. This type of failure can, and probably will, be viewed as stigmatizing in the same manner as a positive biochemical drug test. In fact, here the unknown nature of the RTP test result may work against itself. A worker could be found to have a positive biochemical drug test that could easily and empirically be attributed to a prescription medication. With a "failure" on the RTP test, the employee's state of unpreparedness is not defined and remains open to speculation by management, and perhaps other employee rumors as well.

> Recommendation. Specific detailed contingencies must be implemented along with an RTP test program to deal with employees who do not perform consistently on the RTP test. This type of program should provide support and effective secondary investigative procedures that are well understood by the employee (i.e., follow-up biochemical drug screening, counseling, etc.)

17. A Final Note on Issues and Problems in RTP Testing

Throughout this discussion of RTP testing there has been an attempt to isolate a large number of issues that, when entangled, make it hard to understand the true nature of the RTP concept. It should be recognized that when actually implemented in an operational environment, all of these issues must be reintegrated into a functional testing system. Many of the issues raised must be prioritized in importance with respect to the specific testing situation at hand, and undoubtedly many will be compromised. Ultimately, what seems most important are some of the questions that opened this section. How is one defining RTP? What is to be predicted by the RTP test? Is the RTP test valid?

IMPLEMENTATION PROBLEMS OF RTP MEASURES

1. Testing Methodology

The implementation of an RTP test program requires consideration of a number of methodological issues beyond selection of the test itself. One issue is the frequency of testing. Although vendors may suggest that all employees need not be tested on a daily basis for most jobs, at least one vendor states that an ideal use for their product is in the *daily* screening of all employees in safety-sensitive and other critical positions. An extension of this approach might be testing more than once per day, for example, following breaks or lunch prior to returning to work. This would be even more important in the case of an extended work day to detect the presence of accumulated fatigue. Another implementation issue is the time of testing during the day, particularly for those workers on rotating shifts.

The impact on stability of the RTP measure of the time interval between testing and the time of day for testing must be determined prior to implementing any test schedule. If a worker is not tested on Friday, is off for the weekend, and perhaps not tested the following Monday, the four-day test gap may influence RTP performance, either resulting in test "failure" or lack of a stable baseline. The effect is potentially more extreme with random, once-a-week testing. Here, it is conceivable that testing may occur on Monday of the first week and not until Friday of the second week, resulting in a 10-day test-retest interval. The use of "warm-up" trials to moderate this effect has been suggested, but this practice must also be approached with caution. It is not clear how many warm-up trials should be allowed or how they should be included in the ongoing establishment of the individual baseline.

Another methodological issue in RTP testing is whether to use a single-shot screening approach or a repeated measures application. Most testing situations allow for multiple testing to build a self-referenced comparative baseline. This appears to be the best approach at present. It is feasible that this approach can also reveal long-term trends in cognitive processing ability as a result of aging or the onset of disease. However, some RTP tests may be more vulnerable to reduced reliability or validity through repeated testing. Related to the issue of testing frequency, this raises additional questions of reactivity of measures and reliability in particular. How often *should* one administer the RTP test to workers? How long can one go without administering the RTP test and still retain trained RTP performance in the worker? A related issue is how changes in one's work shift might influence performance on the RTP test.

Situations that will not allow repeated testing to build a within-subject comparative data base, may have to adopt a single-shot approach utilizing more general normative data for comparisons. This appears to be less desirable than a within-subject approach. If a single-shot approach is needed, the actual normative data used for a comparative base is obviously of great concern. The data should be linked directly to the type of subject population undergoing RTP testing.

In a similar vein, RTP testing criteria "by job classification" seems to be an important issue. If testing pilots (or Air Traffic Control specialists), and if the RTP test has predictive validity for job performance, then care should be taken to ensure that pilots are compared to pilots in any between-subjects comparisons that are performed. The unique skills and abilities that may be needed for any specific job might be quite different even across jobs that share considerable content with one another.

The type of test stimuli for RTP testing may be very important. Are the stimuli appropriate for age, ethnic group, gender, and social status? Do the stimuli vary from day to day? If they do, is it possible that on one day the stimuli could be significantly harder or easier than on other days?

Employee motivation seems to be an important issue to some concerned with RTP testing. Some have suggested that the RTP task should not be boring or the subjects may begin to do poorly (false negatives), or fail to comply. They suggest more exciting tasks, game-type tasks, or tasks with high face validity for the job. In one way this may be a "non-issue." While it might be possible that some workers could find a simple task boring, the constant thought that their job and income depended upon their performance would probably address motivation. The one real advantage that more complex tasks might provide is the embodiment of the more sophisticated type of skills and abilities that are required in actual work. It is the inclusion of these more sophisticated abilities that may make the tests more sensitive to the effects of risk factors as well. On the other hand, workers who are not motivated for their work, but who recruit motivation to pass the RTP test, may be at risk later on the job as their motivation wanes. This may be another disguised problem in RTP testing -- the fact that an RTP test is designed to test state conditions also means that it is vulnerable to other states, such as recruited motivation and the temporary recruitment of skills or abilities.

In assessing RTP testing in general, one question ought to be asked and empirically verified at some point. Does RTP testing with a valid test improve employee compliance to rules, safety, and performance over the mere implementation of RTP testing -- regardless of the test used? In other words, because there is little real validity data, perhaps the claims of RTP testing are really a form of placebo effect, or more accurately, a form of the Hawthorne effect. As in the classic Hawthorne study, simply because one institutes testing, and because the testing looks "official" and believable, perhaps the employees change behavior in response to the *change* itself and not the technical aspects of the program. Thus, behavioral changes occur not as a result of a valid RTP test program, but rather as a simple function of someone having raised the issue and having made a change.

2. Risk Assessment in RTP Testing

Ultimately, risk assessment must play some role in RTP testing. Risk assessment is the process whereby the employer must make decisions regarding the degree of risk that can be taken given the potential cost to property, employees, the business, and the public. This requires sophisticated trade-off decisions weighing numerous factors. More specific to RTP testing, the employer must decide what degree of tolerance in RTP test performance can be accepted. Setting performance standards too high would result in needlessly disbarring workers from their jobs. Setting standards too low would allow impaired workers on the job and would significantly raise the potential for job-related accidents. If the cost of errors in RTP testing in terms of accidents is very low, while the potential payoff of keeping slightly impaired workers on the job is high, do we act more leniently? In other words, could there be times when certain people found to be "not ready to perform" are acceptable for work because the risk or the cost of failure is so low? Because most standards for RTP testing are set by the vendor, the consumer can be totally excluded from the risk assessment decision.

3. Test Length

Are brief testing samples sufficient for assessing RTP? Can we determine RTP in one three-minute to five-minute trial? Some of our recent UTC-PAB STRES Battery research (Schlegel and Gilliland, 1992) explored extended performance effects and revealed considerable variation between initial trials and subsequent trials. Of course, these results are based on a limited number of trials and a limited number of tasks and dependent measures. However, these data do raise serious questions about the relationship between brief task samples, during which the person may recruit maximal effort that could never be maintained over a longer period, and more extended job performance. Perhaps one of the reasons that brief behavioral tests have had difficulty in predicting job performance well is they may tap the resources used in the job, but not as they are applied to the job. When employees take a brief test they may apply all their available effort to perform maximally. When employees are on the job, they may pace themselves with the estimation of what remains of the work day on their mind, as well as what level of energy and resources they think they can sustain over that period. Perhaps RTP tests could be even more effective if they avoided the type of behavior that is fortified by the heavy recruitment of resources for a very short period of time.

In this same regard, is the use of a second trial in the case of a "failure" situation sufficient to assess a consistent "failure." Or, should multiple trials, or perhaps a different test, be used?

4. Need for Cross-Validation

The validity for RTP testing is sketchy at best. Certainly, there is a formidable literature demonstrating drug, alcohol, and stress effects on human performance. However, a large leap in logic is taken between this literature and the application of a single, brief test to predict the presence of risk factors. What is needed is research that verifies that brief testing can consistently provide evidence of risk factor effects.

Such a demand is not unlike the demand for cross-validation studies of other types of test instruments. These types of studies are performed fairly easily in a laboratory -- much less easily in the field. Once criterion groups (for example, normal subjects and subjects exposed to some risk factor in a double-blind procedure) are clearly defined, then it can be determined which RTP candidate tests will differentiate between the groups. Any of the "reactive" candidate RTP tests could constitute the actual RTP test. But, before we can assume it is valid we need to cross-validate by testing another sample of people to determine whether we can actually identify among them a random number of individuals exposed to risk factors (administered once again in a double-blind paradigm).

Through the existing literature, one could presumably identify candidate RTP tasks -- the first step. What is missing is the research that verifies that these tasks are valid measures of risk variables. Without such cross-validation studies we are left with rational justifications and simple criterion validity studies only. Without such cross validation studies, the prudent approach to implementing RTP field testing ought to be a very cautious one.

A requirement for cross-validation studies at this point in the state-of-the-art might be stringent, but RTP tests should eventually stand before this test. Vendors should be able to provide such data. It is perhaps the easiest type of data to collect, aside from seeking support in existing literature bases.

5. RTP Testing with Restricted Range, General, and Special Samples

To what degree is RTP robust to factors such as aging? Aging often brings on the same type of decrement as seen with some drugs, i.e., diminishing or loss of memory, less psychomotor skill, etc. How can aging workers be protected in RTP testing? Certainly the use of each subject's own performance means will address part of this problem. But, will we need aging appropriate (gender appropriate, etc.) scaling of test scores as well? On the other hand, there is some indication in the literature (e.g., Collins & Mertens, 1988) that older individuals may be more sensitive to alcohol. This interaction between age and stressor effects may modify the validity of the RTP test. Also, when is a gradual decline in RTP performance due to aging significant enough for concern?

6. Hidden Costs in RTP Testing

No RTP measure can differentiate specific risk variables, i.e., differentiate well (or perhaps, at all) between drugs or alcohol or fatigue, for example. At best, RTP measures detect the lack of performance capacity at the moment. Of course, the lack of performance capacity is usually assumed to be caused by some risk factor. This is both an advantage and disadvantage of RTP testing. In a nonspecific manner, the RTP test detects acceptable or non-acceptable performance capacity, but does not define the cause. This requires that for RTP testing to be effective, there must be other mechanisms put in place to assess "rejection" cases. RTP testing is not a simple case of instituting "black boxes" and training people in simple testing procedures. RTP testing, it would seem, requires additional administrative overhead including a system to further assess any cases of "rejection." This system, which may be elaborate and involve one-on-one counseling, is an added cost of RTP testing that may not be included in what appears to be the overt costs of an RTP test program, i.e., it may be a hidden cost.

Employee time, perhaps much of it on a continuing basis, to assess and negotiate the implementation of the RTP testing system in the workplace could be viewed as a hidden cost. Any program of worker screening is going to require considerable labor-management negotiation procedures and agreements. These proceedings take workers off their jobs and produce a hidden cost. The nature of RTP testing is not as clear-cut as biochemical screening, and could feasibly require even more such labor-management negotiations, especially considering the implications of a "failure" result. In other words, what constitutes a failure? What is done after a failure? How is the employee reassigned? For how long are they reassigned? Many of these issues are complicated with RTP testing because no one knows why there was a failure to begin with. Record keeping will also be labor intensive. Many employers may have to hire additional staff to manage the RTP test records.

Designating someone as "not ready to perform" is one issue. Designating them at a later date as "ready to perform" is another. What procedures does one follow in both cases? To what degree do labor and management representatives interact on this issue? Who sets the standards of performance? What price in terms of time and employee "downtime" is involved here?

The simple implementation of an RTP test program appears on the surface to be less expensive per person than biochemical testing. But, let's examine the basis for comparison. Biochemical drug screens are very expensive per screen. However, not every employee is screened every day. In many cases, only a small percentage of the work force is subjected to random screening at any given time, and the threat of screening appears to be one of the most powerful deterrents at work. In this case, the actual cost of biochemical screens could be contained. If one uses RTP testing, employees will typically be tested more frequently (but perhaps not every day) to ensure maintenance of baseline performance on the RTP test. Now, one could argue that even at a fee of $200 per year, that is less than random drug screens for that employee. Yes, it probably is. But, do drug screens require a total of 5 to 15 minutes of every workday or two to three work days per week? That time period, taken out of the workday of every employee, represents a substantial cost, which is now essentially overhead (read, "reduction in profit margin").

When an employee is found to "fail" an RTP test, what do they do? One solution is reassignment to non-safety sensitive jobs. That sounds good in theory, strategy ended up being exactly the opposite.) In this sense, there is considerable room in RTP testing for what Cook and Campbell (1976) refer to as "reactivity of measures." That is, our measurement technique itself introduces effects we had not expected. Another example of unexpected outcomes is the possible development by third parties of "home versions" of RTP tasks. Computer "hackers" within an employment setting could easily reproduce many of the tasks being used as RTP tests. Employees could then practice at home, distorting or grooming performance for test sessions.

8. The Case of "Falsing"

It is always possible that employees may purposefully attempt to manipulate the RTP test to influence the testing outcome, i.e., "falsing." Procedures must be established to guard against falsing. This may require greater vigilance during testing and greater attention and creativity in test result scoring.

9. Need for RTP Testing Standards

Closely related to such questions as the need for cross-validation studies is a more general demand for testing standards. The whole domain of RTP testing presents a somewhat unique situation for industry. On the one hand is testing technology that is built on what appears to be fairly secure scientific grounds (i.e., past research on risk factor effects on human performance). On the other hand, RTP tests are being promoted and sold often with very little evidence for the effectiveness of the specific RTP test in question.

Some have suggested that RTP testing must seek both the testing tools and the testing standards (Elsmore, personal communication). It appears that there are some RTP vendors who are offering the testing tools, but it does not appear that there are testing standards yet. By "testing standards," it is meant clear procedures, norms, validity studies, etc. that should be available prior to the marketing of an RTP test. The field of RTP testing is moving rapidly enough, the issues surrounding RTP testing are important enough (i.e., drug screening), and the cost is certainly high enough, to warrant some degree of concern from the vantage point of the consumer. This is not an unreasonable demand. For example, the American Psychological Association, as well as a number of other professional associations, have standards for the vending of psychometric tests.

REFERENCES

Backer, T. E. (1987). *Strategic planning for workplace drug abuse programs.* NIDA Publication ADM-87-1538, Washington, DC: U.S. Government Printing Office.

Blower, D., and Dolgin, D. (1991). *An evaluation of performance based tests to improve Naval Aviation selection* (Tech. Report NAMRL-1363). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Bureau of National Affairs. (1986). *Alcohol and drugs in the workplace: Costs, controls, and controversies.* Rockville, MD: The Bureau of National Affairs, Inc.

Bureau of Statistics, United States Department of Labor. (1989). *Survey of employer anti-drug programs* (Report 760). Washington, DC: U.S. Government Printing Office.

Collins, W. E., and Mertens, H. W. (1988). Age, alcohol, and simulated altitude: Effects on performance and breathalyzer scores. *Aviation, Space, and Environmental Medicine, 59,* 1026-1033.

Cook, T. D., and Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial /organizational psychology.* Chicago, IL: Rand McNally.

Cronbach, L. J. (1970). *Essentials of Psychological Testing.* New York: Harper & Row.

Crouch, D. J., Webb, D. O., Buller, P. F., and Rollins, D. E. (1989). A critical evaluation of the Utah Power and Light Company's substance abuse program: Absenteeism, accidents and costs. In S.W. Gust and J. M. Walsh (Eds.), *Drugs in the workplace: Research and evaluation data.* (Research Monograph 91, pp. 169-193). Washington, DC: National Institute on Drug Abuse.

Greenfield, P. A. (1989). Drug testing and the law. *Employee Responsibilities and Rights Journal,* 2, 11-26.

Greenfield, P. A., Karren, R. J., and Giacobbi, J. K. (1989). Drug testing in the workplace: An overview of legal and philosophical issues. *Employee Responsibilities and Rights Journal,* 2, 1-10.

Guthrie, J. P., and Olian, J. D. (1989, April). Drug and alcohol testing programs: The influences of organizational context and objectives. Paper presented at the Fourth Annual Conference of the Society for Industrial/Organizational Psychology, Boston.

Hamilton, J. (1991). A video game that tells if employees are fit for work. *Business Week,* June 3.

Hanson, A. (1990). What employees say about drug testing. *Personnel, 67,* 33-36.

Hartstein, B. A. (1987). Drug testing in the workplace: A primer for employers. *Employee Relations Law Journal, 12,* 577-604.

Karren, J. R. (1989). An analysis of the drug testing decision. *Employee Responsibilities and Rights Journal, 2,* 27-37.

Lumsden, H. (1967). The plant visit: A critical area of recruiting. *Journal of College Placement, 27,* 74-84.

Maltby, L. L. (1990). Put performance to the test. *Personnel, 67,* 30-31.

Murphy, K. R., Thornton, G. C., and Prue, K. (1991). Influence of job characteristics on acceptability of employee drug testing. *Journal of Applied Psychology, 76,* 447-453.

Sanders, A. L. (1989, April). A boost for drug testing. *Time,* 62.

Schlegel, R. E., and Gilliland, K. (1992). *Development of the UTC-PAB normative database* (Tech. Report AL-TR-92-). Wright-Patterson AFB, OH: Armstrong Laboratory.

Seeber, R. L., and Lehman, M. (1989). The union response to employer-initiated drug testing programs. *Employee Responsibilities and Rights Journal,* 2, 39-48.

Sitomer, C. J. (1989, March). When drug testing is justified. *Christian Science Monitor,* 13.

Stone, D. L., and Kotch, D. A. (1989). Individual's attitudes toward organizational drug testing policies and practices. *Journal of Applied Psychology, 74,* 518-524.

Stone, D. L., and Vine, P. L. (1989, April). Some procedural determinants of attitudes toward drug testing. Paper presented at the Fourth Annual Conference of the Society for Industrial/Organizational Psychology, Boston.

Thorson, H., and Thomas, W. (1968). Student opinions of the placement process. *Journal of College Placement, 29,* 82-84.

Wiggins, J. S. (1973). *Personality and Prediction: Principles of Personality Assessment.* Reading, Massachusetts:Addison-Wesley Publishing Company.

Wrich, J. T. (1988, January/February). Beyond testing: Coping with drugs at work. *Harvard Business Review,* 120-128.

Appendix A

Review of Computer-Based Performance Assessment Batteries

The availability of modestly-priced microcomputers has encouraged the development of numerous tests and test batteries for assessing cognitive performance and the effects of various stressors. Several of the individual tests are historically founded in traditional pencil-and-paper tests of cognitive ability. Others take advantage of unique capabilities afforded by a computer-based test, such as millisecond response timing, dynamic movement for tracking and monitoring tasks, and the simultaneous presentation of multiple tasks to examine attention and time-sharing resources.

The following review provides overviews of many of the current and popular performance task batteries, descriptions for many of which are not available in any published source. Special attention was given to include those batteries most likely to provide candidate RTP measures. The review is not exhaustive but is intended to provide readers with a representative sample of available batteries. Many of these batteries are in development or have been recently released. For that reason, very little research has been conducted with them, and in some cases, normative data are not even available. Where possible, the authors have included information they have received through unpublished manuscripts, personal communications, and personal contacts. Table A-1 provides a cross-listing of tasks across batteries to aid the reader in comparing the various batteries.

Another review of computer-based tests that are used for neuropsychological and performance-based assessment was provided by Kane and Kay (1992). In their review, thirteen major computer-based cognitive performance assessment batteries were examined with information provided on (1) development history, (2) hardware requirements, (3) included tasks, (4) test administration, (5) parameter options, (6) data output, (7) norms, and (8) validation studies. Information is also provided on individual tests common to several batteries. The following taxonomy was used to classify the individual tests: Simple Motor Tests, Reaction Time Tests, Attention-Concentration Working Memory, Learning and Memory, Spatial Perception/Reasoning, Calculations, Language, Complex Problem Solving, Dual-Tasking and Multi-Tasking.

Table A-1. Test by Battery Cross-Reference List.

Battery	ACS	AGARD STRES	ANAM	APTS	B-MAPS	CCAB	COG-SCREEN	CTS	MATB	MPTB	NMRI-PAB	SYNWORK	UTC-PAB	WRPAB
Motor														
Finger Tapping				***										
Visual Motor Tracking		***	***				***	***	***				***	
Coordination/Steadiness							***							
Reaction Time														
Simple Reaction Time	***			***					***	***				
Complex Reaction Time	***	***	***	***								***	***	***
Attention/Working Memory														
Continuous Performance	***			***				***					***	
Neisser Task				***							***		***	***
Symbol-Digit/Code Subst.				***	***		***						***	***
Sequence Comparison	***			***			***						***	
Sternberg Search/Scan		***	***	***				***			***	***	***	***
Stroop				***							***		***	***
Digit Recall	***						***							
Number Recall/Prev. Numb.				***			***							
Running Memory			***											
Switch/Shift Attention							***							
Time Estimation				***				***					***	***
Visual Monitoring							***	***	***	***		***	***	
Auditory Sequence Compare							***							
Dichotic Listening													***	
Learning and Memory														
Match/Non-Match to Sample							***				***		***	***
Pattern Comparison	***			***	***						***		***	***
Associate Learning				***			***							***
Sequence Memory				***									***	***
List Recall/Recognition	***				***									
Text Memory	***													
Spatial Rotation-Sequential		***	***	***	***			***					***	
Repeated Acquisition											***		***	
Spatial Perception/Reasoning														
Pattern Matching				***	***					***			***	
Manikin				***	***		***				***		***	***
Spatial Rotation-Simul.			***	***										
Clock Faces	***													
Spatial Visualization	***			***										

Table A-1. Test by Battery Cross-Reference List.

Battery	ACS	AGARD STRES	ANAM	APTS	B-MAPS	CCAB	COG-SCREEN	CTS	MATB	MTPB	NMRI-PAB	SYNWORK	UTC-PAB	WRPAB
Calculations														
Arithmetic Computation	***									***		***	***	***
Math Reasoning							***							
Serial Addition/Subtraction		***	***	***				***					***	***
Language														
Linguistic Processing				***				***					***	
Verbal Analogies	***													
Grammatical/Logical Reas.		***	***	***		***		***			***		***	***
Word Anagrams						***								
Complex Problem Solving														
Object Match	***													
Rule Discovery							***							
Tower Puzzle						***								
Route Planning						***								
Category Sorting					***									
Following Directions						***								
Information Purchase						***								
Series Completion						***								
Novel Sequencing							***							
Encoding/Decoding														***
Code-Lock										***				
Process Control/Scheduling									***					
Dual-Tasking/Multi-Tasking		***				***	***		***			***	***	
Sternberg/Tracking		***											***	
Visual Monitoring/Sequence							***							
Tracking/Previous Number							***							
Numbers and Words						***								
Mark Numbers						***								
Multiple Combinations										***				

Multiple Task Performance Battery (MTPB)

Although originally developed using mechanical components and hardwired logic, the MTPB represents an early implementation of a sophisticated multiple-task performance assessment tool (Chiles, Alluisi, and Adams, 1968). Developed at the Lockheed-Georgia Company and originally used by researchers from the USAF Aerospace Medical Research Laboratories during the late 1950's and 1960's, the MTPB provided assessment of monitoring, arithmetic, and complex code-solving performance in a time-sharing work environment. In addition to providing a model for later multiple-task tests, such as the Synthetic Work Task (SYNWORK) and the NASA Multi-Attribute Task Battery (MATB), individual tests have been drawn from the MTPB. For example, the Probability Monitoring task of the Criterion Task Set (CTS) was modeled after a similar task in the MTPB. A computer-based version of the MTPB has been developed and is being used by researchers at the FAA Civil Aeromedical Institute.

Reported reliabilities for performance measures on the original MTPB are in the range of 0.70 to 0.97. The MTPB has been used to (1) evaluate performance during long periods of confinement (Chiles, et al., 1968), (2) to assess the effects of alcohol (Chiles and Jennings, 1970), altitude and high temperature (Chiles, Iampietro, and Higgins, 1972), and (3) as a performance predictor for air traffic controller trainees (Chiles, Jennings, and West, 1972). Its potential as an RTP tool lies in its ability to present a complex cognitive task involving attention time-sharing.

Automated Portable Test System (APTS)

Based on the Navy's Performance Evaluation Tests for Environmental Research (PETER) program initiated in the late 1970's, the Automated Portable Test System (APTS) presents 21 tasks on a portable computer (Bittner, Smith, Kennedy, Staley, and Harbeson, 1985). A key feature of the tests selected for inclusion is their high degree of stability and accompanying suitability for repeated administrations. Stability in this case refers to the ability to rapidly reach asymptotic mean performance levels, with constant variance and high differential stability across subjects within a group.

The PETER program initially reviewed over 150 tests, primarily in paper-and-pencil form, and consisting of classic cognitive psychology abilities tests. Tests offering high stability were selected for computer implementation and the reliabilities of the computer-based versions were verified (Bittner, Carter, Kennedy, Harbeson, and Krause, 1986).

Merkle, Kennedy, Smith, and Johnson (1985) and Kennedy, Dunlap and Kuntz (1989) provide reviews of various studies using the APTS to assess behavioral effects of stressors including drugs used to treat motion sickness, alcohol (Kennedy, Wilkes, and Rugotzke, 1989), altitude (Kennedy, Dunlap, and Kuntz, 1989), and chemoradiotherapy related to bone marrow transplantation (Parth, Dunlap, Kennedy, Lane, and Ordy, 1989). A major advantage of the APTS with respect to

RTP testing is its suitability for repeated measurement applications due to the high stability of the tests. The Delta™ RTP system is based on the APTS.

Criterion Task Set (CTS)

The CTS, developed by Shingledecker (1984) at the USAF Aerospace Medical Research Laboratory, represents one of the earliest instances of computer-based performance assessment. Although it was designed to provide a set of standardized loading tasks to evaluate the relative sensitivity, reliability, and intrusiveness of a variety of available workload measures, the CTS has been used directly for performance assessment. One of its major features is the fact that it is based on current multiple resource theories of information processing (Wickens, 1992) and provides tasks that tap various stage, code, and mental activity resources. The nine CTS tasks include Display Monitoring, Unstable Tracking, Interval Production, Continuous Recognition, Grammatical Reasoning, Linguistic Processing, Mathematical Processing, Memory Search, and Spatial Processing. A noted advantage of the CTS is that eight of the tasks were designed to provide three distinct levels of difficulty (representing three different levels of mental workload).

Although implementation of the CTS on the Commodore 64 computer system formerly represented an advantage in terms of system cost and response timing capability, the obsolescence of these systems and the cost reduction for PC compatibles makes the Commodore version of the CTS currently less viable. However, tasks from the CTS are included in the UTC-PAB and other PC-based batteries (but usually only at one difficulty level). Payne, Pike, and Birkmire (1992) have implemented most of the CTS tasks on PC compatible equipment.

Schlegel and Gilliland (1990) evaluated the CTS and provided normative data based on 123 subjects. Depending on the task, two-day test-retest reliabilities ranged from 0.59 to 0.91 for response time measures. A cluster analysis of the database was performed to study the construct validity of the CTS and its relatedness to multiple resource theory. Four distinct clusters were identified, leading the authors to conclude that the CTS did represent a battery of tasks tapping separate information processing resources and stages. Schlegel and Gilliland also examined the sensitivity of the CTS to noise stress, sleep deprivation, and caffeine and found an overall lack of effect for most tasks at the stressor levels employed. The developers of NovaScan™ were also the originators of the CTS, upon which the STRES and much of the UTC-PAB are based.

Walter Reed Performance Assessment Battery (WRPAB)

The WRPAB was designed as a research tool for assessing performance changes over time, treatments, or dosages (Thorne, Genser, Sing, and Hegge, 1985). As an early computer-based test battery, the WRPAB has inspired much of the development and design of subsequent PABs such as the UTC-PAB, AGARD-STRES, and COGSCREEN. Although originally written for Apple II computers, it currently runs on PC compatibles. The WRPAB currently consists of 22 tasks.

The WRPAB has been used to investigate circadian rhythms, sleep deprivation, fatigue, physical conditioning, bright light, hypoxia, heat stress, sickle cell anemia, HIV, and a host of drugs. Sensitivity has been demonstrated for atropine, amphetamine, antihistamine, and fatigue.

Studies demonstrating the WRPAB's sensitivity to the effects of drugs, alcohol, and environmental stressors have been conducted by the Addiction Research Center (Higgins, Lamb, and Henningfield, 1989). Reeves (1990) found a relationship between antihistamine dose and performance on selected WRPAB measures.

The fact that WRPAB tests were designed for repeated measures applications makes them attractive candidates for use as RTP measures.

Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB)

The UTC-PAB was developed by the Tri-Service Joint Working Group on Drug-Dependent Degradation of Military Performance (JWGD^3 MILPERF), now organized as the Office of Military Performance Assessment Technology (OMPAT). Originally consisting of a collection of 25 computer-based tests, the UTC-PAB allowed researchers to select a subset of tests to configure a desired performance assessment battery while maintaining standardized hardware, software and procedures (Hegge, Reeves, Poole, and Thorne, 1985; Englund, Reeves, Shingledecker, Thorne, Wilson, and Hegge, 1987).

UTC-PAB tests were selected from various sources including the Criterion Task Set, the Walter Reed PAB, and the PETER battery. Current versions of the software run on PC compatible hardware and do not require any additional boards or response apparatus. One existing subset of the battery is configured as the NATO AGARD-STRES Battery. Schlegel and Gilliland (1992) provided normative data and reliability characteristics for this subset of tasks and also demonstrated correspondence of performance with similar tasks from the Criterion Task Set.

Nesthus, Schiflett, Eddy, and Whitmore (1992) evaluated the sensitivity of a subset of UTC-PAB tests to terfenedrine (not sensitive) and diphenhydramine (sensitive) during sustained operations. Other studies have demonstrated test sensitivity to hypoxia, amphetamines, alcohol, and temperature changes. A major advantage of the UTC-PAB is that many of the individual tests have a long history of use in experimental psychology.

Naval Medical Research Institute Performance Assessment Battery (NMRI-PAB)

The NMRI-PAB was developed to assess operational environment effects on military performance (Schrot and Thomas, 1988) using eight standardized tests of response accuracy, logical reasoning, response acquisition, short-term memory, attention, spatial orientation, pattern matching, and color and form discrimination. The BASIC software runs on PC compatibles with EGA video and requires an additional timing board.

In their FAA-sponsored report of comparative studies of cognitive tests, Horst and Kay (1988) presented normative data for normal healthy pilots on three of the tests. Various tests within the NMRI-PAB have been shown to be sensitive to the effects of environmental stressors such as underwater diving in hot water (Thomas, Schrot, Ahlers, Thornton, Dutka, Armstrong, Kowalski, and Shurtleff, 1991), cold water (Doubt, Weinberg, Hesslink, and Ahlers, 1989) and to antihistamines (Schrot, Thomas, and Van Orden, 1990).

Advisory Group for Aerospace Research and Development Standardized Tests for Research with Environmental Stressors (AGARD-STRES)

The NATO AGARD-STRES battery (Santucci, Farmer, Grissett, Wetherell, Boer, Gotters, Schwartz, and Wilson, 1989) comprises a subset of seven UTC-PAB tests with rigidly fixed test parameters. The battery represents an attempt at international standardization of computer-based cognitive tests for use in environmental stress and performance research. The full battery requires approximately 30 minutes. A PC compatible version was implemented by Reeves, Winter, LaCour, Raynsford, Vogel, and Grissett (1991). Although the AGARD-STRES battery was developed as a baseline battery for repeated measures testing, the previously described normative database and stressor sensitivity effects are applicable (Schlegel and Gilliland, 1992).

Automated Neuropsychological Assessment Metrics (ANAM)

ANAM incorporates five of the AGARD-STRES tests in a format suitable for clinical neurological screening (Reeves, Winter, LaCour, Raynsford, Kay, Elsmore, and Hegge, 1992). As with AGARD-STRES, the ANAM is particularly well-suited for both comparative assessment with respect to norms and for repeated measures applications. In contrast to AGARD-STRES, the ANAM battery allows the examiner to alter various test parameters to meet different assessment needs. Parameter modifications include the specification of the response device (keyboard vs. mouse), test duration, interstimulus interval, stimulus presentation time, presentation of instructions, and elements of stimulus sets.

Previously described normative data and stressor sensitivity studies for AGARD-STRES (Schlegel and Gilliland, 1992) are equally applicable to the corresponding ANAM tests.

Assessment of Cognitive Skills Battery (ACS)

The ACS was developed by a mixed panel of neuropsychological, clinical, testing, and statistical specialists to assess long-term cognitive status changes in physicians and other professionals (Powell, Catlin, Funkenstein, Kaplan, Ware, Weintraub, and Whitla, 1990). The ACS runs on PC compatible hardware and consists of thirteen tests. A normative database was established using more than 1100 volunteer physicians (90% male).

Internal consistency reliability coefficients ranged from 0.87 to 0.98 for the various tests. Test-retest reliability was low (0.14 to 0.75, mean of 0.43) due perhaps to restrictions in range, a small number of items per subtest and item difficulty in relation to age. Concurrent validity was evaluated by classifying subjects as normal or impaired based on the ACS and a composite score from three traditional neuropsychological measures (Wechsler Memory Scale - Revised, Boston Naming Test, Wisconsin Card Sorting Test). The classifications were in agreement for 77.6% of the cases.

Bexley-Maudsley Automated Psychological Screening (B-MAPS) and Category Sorting Test

The Bexley-Maudsley Automated Psychological Screening was developed by Acker and Acker (1982) as a time-efficient, cost-effective screen for alcoholics and individuals with subtle forms of cognitive impairment. It consists of six subtests and requires 45 minutes to administer and score. Although originally implemented on Commodore PET 32k and Apple II computers, PC and MacIntosh versions are currently under development.

The battery's primary use has been in studies involving alcoholics, although published data on the battery's effectiveness in assessing cognitive impairment and its relationship to other measures is limited. Glenn and Parsons (1990, 1991) demonstrated that the Manikin, Pattern Comparison, and Category Sorting tests were sensitive to cognitive impairments in a group of relatively young female alcoholics ages 21-49. They used efficiency scores based on the ratio of number correct to total time spent on task, having found no significant differences between alcoholics and controls when using only the number of correct responses.

COGSCREEN™

COGSCREEN was developed for the Federal Aviation Administration as a screening instrument for detecting changes in the cognitive functioning of aviators which might result in poor pilot judgment or slow reaction time in critical situations (Horst and Kay, 1991). COGSCREEN consists of eleven tests including tests of memory, mathematical reasoning, spatial processing, divided attention, shifted attention, and tracking. Because it is primarily a screening test, COGSCREEN has undergone extensive normative data development involving commercial and military aviators, distinguished by age groups. The battery is also being administered to Russian pilots following the same testing protocol.

Construct validity was evaluated by comparing COGSCREEN performance with performance on corresponding pencil-and-paper tests. Correlation coefficients ranged from 0.44 for the Number Pathfinder test to 0.80 for the Symbol Digit Coding test.

One of COGSCREEN's unique features is the use of a light pen for response input. This reduces the negative impact of minimal prior computer and keyboard experience as a confounding factor in assessing subject performance.

Complex Cognitive Assessment Battery (CCAB)

The CCAB represents a departure from the previously presented batteries in that it was designed to evaluate performance on tasks that require high-level, complex cognitive skills. The battery was initially developed to provide a tool for evaluating the cognitive performance effects of prophylactic drugs used in chemical defense (Samet, Geiselman, Zajaczkowski, and Marshall-Miles, 1986). The battery runs on PC compatible hardware.

A "top-down" design approach was used. This approach started with the formulation of constructs to be measured and led to the design of specific tests to measure the constructs. Existing performance taxonomies were used to derive fourteen complex cognitive constructs: attention to detail, perception of form, memory retrieval, time sharing, comprehension, concept formation, verbal reasoning, quantitative analysis, planning, situational assessment, decision making, communication, problem solving, and creativity. Nine tests were developed to measure various combinations of the constructs.

Normative data have been provided by Geiselman and Samet (1986) and by Kay and Horst (1988), and a large set of normative data is being generated in association with pilot selection research at the Army Research Institute at Fort Rucker, Alabama. Reported two-day test-retest reliability ranges from 0.66 on the Mark Numbers test to 0.95 on the Following Directions test. A weighted average test-retest reliability coefficient across a six-test subset was 0.80 (Geiselman and Samet, 1986). Kay and Horst (1990) found significant practice effects on CCAB tests administered at one-week intervals across four weekly sessions and determined that males were faster and more accurate on the Following Directions and Route Planning tests.

There are few published investigations employing the CCAB despite its popularity among researchers and its somewhat unique task collection. Although the CCAB has been used in pilot selection studies and to examine the effects of sleep apnea, aspartame, and antihistamines, data are available only in technical reports, abstracts, and through personal communications. As with tests in other batteries mentioned previously, CCAB tests have been found to be sensitive to diphenhydramine but not to terfenadine or astemizole.

The CCAB has particular appeal as a potential RTP measure in that the tests appear to tap a level of complex problem solving not addressed in other PABs. The tests are creative and challenging. They tend to have an obvious multifactorial structure, and in most cases involve at least some degree of divided and sustained attention. As in a work environment, the CCAB presents subjects with complex jobs that must be completed within certain deadlines. This multifactor structure and complexity should make it particularly sensitive to stressors affecting high-level complex brain function.

Synthetic Work Task (SYNWORK)

The SYNWORK program was designed as a laboratory performance test that is an intermediate evaluation tool between typical PAB tests and full blown simulators. A key feature of SYNWORK is that four tasks are presented concurrently (each in a different quadrant of the display) and thus the task provides a measure of the subject's time-sharing ability. The task runs on PC compatible computers and requires a mouse for response input.

The tasks within SYNWORK (Sternberg Memory Task, Arithmetic Task, Visual Monitoring, and Auditory Monitoring) were selected to sample functional characteristics of the "real-world." Although not intended to be a simulation of any specific real-world job, subjects (helicopter flight crews, intensive care monitors) have commented that they considered SYNWORK to be a reasonably good simulator of aspects of their jobs.

Subjects enjoy the task and motivation problems have been minimal, likely due to the constant provision of performance feedback during the testing. Performance on the various component tasks has been shown to be sensitive to sleep deprivation (Kane and Kay, 1992).

NASA Multi-Attribute Task Battery (MATB)

The MATB was developed at NASA Langley Research Center to provide a comprehensive behavioral metric for assessing operator performance. The battery is actually designed to be implemented as a complex cognitive task not unlike the Synthetic Work Task. However, the user does have the ability to present any of the tasks singly or all of the tasks simultaneously. The task is structured to approximate an aircrew operations environment. In this regard, the MATB includes a Monitoring task that consists of both a set of response time stimuli and a set of probability monitoring dials, a compensatory Tracking task, a Resource Management task that is presented as a fuel tank management task, and an auditory Communications task. A user-friendly script system provides the experimenter with a high degree of control over the scheduling of task onset and offset. The auditory Communications task and the Resource Management task are unique features of this battery (although the Communications task, requiring a second dedicated PC, is somewhat difficult to implement). Another unique feature is that the MATB can be paused at any time for onscreen presentation of the NASA-TLX subjective workload scale -- a helpful feature for those who want concomitant subjective workload ratings. The MATB runs on a 286 or 386 IBM-compatible PC with EGA graphics and a mouse or joystick.

An initial study has been completed that provides baseline data for the MATB (Arnegard, 1990) as well as a contractor report describing the use of the MATB in a study of operator strategy (Arnegard, 1991). Current work is under way at the Human Performance Branch of Armstrong Laboratory at Wright-Patterson AFB and at the Personality Research Laboratory at the University of Oklahoma using the MATB as a complex cognitive task in explorations of cognitive psychophysiology, sustained operations, and stress/adaptation.

REFERENCES

Acker, W., and Acker, C. (1982). *Bexley Maudsley Automated Psychological Screening and Bexley Maudsley Category Sorting Test: Manual.* Windsor, Great Britain: NFER-Nelson.

Arnegard, R. J. (1990). Multi-Attribute Task Battery: The bottom line. Poster presented at the 34th Annual Meeting of the Human Factors Society, Orlando, FL.

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

Bittner, A. C., Jr., Carter, R. C., Kennedy, R. S., Harbeson, M. M., and Krause, M. (1986). Performance evaluation tests for environmental research (PETER): Evaluation of 114 measures. *Perceptual and Motor Skills, 63,* 683-708.

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.

Chiles, W. D., and Jennings, A. E. (1970). Effects of alcohol on complex performance. *Human Factors, 12,* 605-612.

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., Iampietro, P. F., and Higgins, E. A. (1972). Combined effects of altitude and high temperature on complex performance. *Human Factors, 14,* 161-172.

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.

Doubt, T. J., Weinberg, R. P., Hesslink, R. L., and Ahlers, S. T. (1989). *Effects of serial wet-dry-wet cold exposure: Thermal balance, physical activity, and cognitive performance* (Tech. Report NMRI-89-35). Bethesda, MD: Navy Medical Research Institute.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., and Hegge, F. W. (1987). *Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) I: Design and specification of the battery.* San Diego, CA: Naval Health Research Center.

Geiselman, R. E., and Samet, M. G. (1986). *Complex Cognitive Assessment Battery (CCAB): Pilot test results* (Report ACC-T-33212). Los Angeles, CA: Analytic Assessments Center.

Glenn, S. W., and Parsons, O. A. (1990). The role of time in neuropsychological performance: Investigation and application in an alcoholic population. *The Clinical Neuropsychologist, 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.

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.

Higgins, S. T., Lamb, R. J., and Henningfield, J. E. (1989). Dose-dependent effects of atropine on behavioral and physiologic responses in humans. *Pharmacology, Biochemistry, and Behavior, 34,* 303-311.

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, OK: Federal Aviation Administration.

Horst, R. L., and Kay, G. G. (1991). *Cognitive function evaluation in medical certification of airmen: Development and validation of a prototype test battery* (Report FAA/933-014-90). Oklahoma City, OK: 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.

Kay, G. G., and Horst, R. L. (1988). *Evaluating cognitive function: A review of mental status tests, neuropsychological procedures, and performance based approaches* (Report on contract DTFA-02-87-C-87069). Oklahoma City, OK: Federal Aviation Administration.

Kay, G. G., and Horst, R. L. (1990). *Effects of aspartame on cognitive performance* (Report on contract DTFA-02-88-C-87069). Oklahoma City, OK: Federal Aviation Administration.

Kennedy, R. S., Dunlap, W. P., and Kuntz, L. A. (1989). *Application of a portable automated performance test battery for the study of drugs and driving performance.* Paper presented at the Second International Symposium on Medicinal Drugs and Driving Performance.

Kennedy, R. S., Wilkes, R. L., and Rugotzke, G. G. (1989). *Cognitive performance deficit regressed on alcohol dosage.* Paper presented at the Eleventh International Conference on Alcohol, Drugs, and Traffic Safety, Chicago, IL.

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, CA.

Nesthus, T. E., Schiflett, S. G., Eddy, D. R., and Whitmore, J. N. (1991). *Comparative effects of antihistamines on aircrew performance of simple and complex tasks under sustained operations* (Technical Report AL-TR-1991-0104). Brooks Air Force Base, TX: Armstrong Laboratory.

Parth, P., Dunlap, W. P., Kennedy, R. S., Lane, L. E., and Ordy, J. M. (1989). Motor and cognitive testing of bone marrow transplant patients after chemoradiotherapy. *Perceptual and Motor Skills, 68,* 1227-1241.

Payne, D. G., Pike, D. A., and Birkmire, D. P. (1992). *Criterion Task Set Workload Assessment Tool.* Poster presented at the Human Factors Society 36th Annual Meeting, Atlanta, GA.

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, MA: Risk Management Foundation of the Harvard Medical Institutions.

Reeves, D. L. (1990). *Assessment of antihistamine and stimulant drug-effects using the Unified Tri-Services Performance Assessment Battery.* Paper presented at the 98th Annual Convention of the American Psychological Association, Boston, MA.

Reeves, D. L., Winter, K., LaCour, S., Raynsford, K., Kay, G., Elsmore, T., and Hegge, F. W. (1992). *Automated Neuropsychological Assessment Metrics documentation: Vol. I Test administration guide* . Silver Spring, MD: Office of Military Performance Assessment Technology.

Reeves, D. L., Winter, K. P., LaCour, S. J., Raynsford, K. M., Vogel, K., and Grissett, J. D. (1991). *The UTC-PAB/AGARD-STRES Battery: User's manual and system documentation* (Tech. Report NAMRL-SR91-1). Washington, DC: Office of Military Performance Assessment Technology, Walter Reed Army Institute of Research.

Samet, M., Geiselman, R. E., Zajaczkowski, F., and Marshall-Miles, J. (1986). *Complex Cognitive Assessment Battery (CCAB): Test descriptions.* Alexandria, VA: U.S. Army Research Institute.

Santucci, G., Farmer, E., Grissett, J., Wetherell, A., Boer, L., Gotters, K., Schwartz, E., and Wilson, G. (1989). *AGARDograph #308, Human performance assessment methods* (ISBN 92-835-0510-7). Seine, France: North Atlantic Treaty Organization Advisory Group for Aerospace Research and Development, Working Group 12.

Schlegel, R. E., and Gilliland, K. (1990). *Evaluation of the Criterion Task Set - Part I: CTS performance and SWAT data - baseline conditions* (AAMRL-TR-90-007). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Schlegel, R. E., and Gilliland, K. (1992). *Development of the UTC-PAB normative database* (Tech. Report AL-TR-92-). Wright-Patterson AFB, OH: Armstrong Laboratory.

Schrot, J., and Thomas, J. (1988). *Performance assessment battery software* (Tech. Report NMRI-88-6). Bethesda, MD: Navy Medical Research Institute.

Schrot, J., Thomas, J. R., and Van Orden, K. S. (1990). *Human performance following antihistamine administration* (Tech. Report NMRI-90-127). Bethesda, MD: Navy Medical Research Institute.

Shingledecker, C. A. (1984). *A task battery for applied human performance assessment research* (Tech. Report AFAMRL-TR-84-071). Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratory.

Thomas, J. R., Schrot, S., Ahlers, S. T., Thornton, M. O., Dutka, A. J., Armstrong, D. W., Kowalski, K. R., and Shurtleff, D. (1991). *Pyridostigmine and warm water diving protocol 90-05: III Cognitive performance assessment* (Tech. Report NMRI-90-97). Bethesda, MD: Navy Medical Research Institute.

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

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd edition). New York: Harper Collins.

Appendix B

Review of the Influence of Selected Risk Factors on Human Performance:

The Research Foundation of Readiness to Perform Measures

Readiness to Perform (RTP) testing has been a natural outgrowth of two simultaneous developments. In the past decade, human performance testing has made dramatic progress in implementing task batteries on microcomputers, freeing researchers from rigid, single-purpose electro-mechanical devices that often made data collection and data reduction cumbersome and time consuming. These microcomputer task batteries provide far greater flexibility in task presentation and more rapid development of new and innovative variations of traditional laboratory tasks. At the same time, screening for drug and alcohol use for high risk occupations was becoming commonplace, yet very costly in terms of time and money. Eventually, the connection between computer-based behavioral tasks that were sensitive to risk factors and the need for less intrusive and more cost-effective drug screening was made, resulting in RTP testing.

One of the major problems facing readiness to perform (RTP) assessment is the lack of research demonstrating validity. Very few RTP measures have undergone actual experimental verification of their usefulness in identifying risk factors. What is far more common is the claim that a specific RTP measure is valid because performance on other similar measures has been shown to be influenced by the experimental introduction of risk factors. For example, an abstract reasoning task might be considered a valid RTP candidate measure because performance on various abstract reasoning tasks has been shown to be degraded when subjects are administered alcohol. In this manner, the intuitive weight of past research can be brought to support the rational selection and use of a specific candidate RTP measure. As noted in the body of this report, no amount of research based on similar measures will provide conclusive evidence for the validity of a specific RTP measure that has not itself undergone carefully conducted validity studies.

Nonetheless, past research on the influence of risk factors on various human performance tasks can aid in selecting those tasks that would be likely candidates for RTP measures. The following is a brief overview of some of the research that demonstrates the sensitivity of various human performance tasks to a variety of risk factors. Given all possible human performance tasks and all possible risk factors, this potential literature base is extremely large. This review will concentrate on only a few of the risk factors that are most often the focus of RTP testing, namely, alcohol, selected drugs, and fatigue. In addition, this review is not intended to be exhaustive. It is merely illustrative of the type of literature that exists in support of the use of human performance tasks as RTP measures.

ALCOHOL

Without question, alcohol is the most abused drug in our society. The negative effects of alcohol nationwide in 1990 were estimated to cost $136 billion. Most of the costs were incurred as a result of lost production, crime, accidents, and treatment. The detrimental influence of alcohol on human performance ability is reflected in the fact that approximately 40% of automobile accidents (and from 10-30% of private aviation accidents) are believed to be alcohol related (see DHHS, 1990; Harper and Albers, 1964; and Modell and Mountz, 1990).

Even more important for the concept of RTP assessment are controlled investigations in field settings that have demonstrated acute alcohol effects on job performance. For example, it has been demonstrated that after alcohol consumption, numerous aircraft piloting skills are degraded, such as radio-signal tracking and air traffic vectoring, observation, and avoidance (Ross and Mundt, 1988). Control of aircraft descent (Ross and Mundt, 1988), stick and pedal control (Tang and Rosenstein, 1967), and numerous other in-flight aircraft control procedures (Billings, Wick, Gerke, and Chase, 1973; Henry, Flueck, and Sanford, 1974) are also degraded by the influence of alcohol. Many of these degrading effects begin to appear at blood alcohol concentrations (BAC) between .03% and .05%. These are well below the level generally designated as legally intoxicated (i.e., 0.1% BAC; see Ross and Mundt, 1988; Ross, Yeazel, and Chau, 1992; Tang and Rosenstein, 1967). Possibly compounding the problem is the suggestion that alcohol absorption rate increases at higher altitudes resulting in a higher BAC than would be experienced at a lower altitude (Higgins, Vaughan, and Funkhouser, 1970). Behavioral affects due to this higher BAC were not demonstrated in this study, however.

Another example of the influence of alcohol on job performance is found in studies of driving behavior. Studies of driving ability under the influence of alcohol, both on closed driving courses and in simulators, generally agree that numerous driving behaviors are affected even at BAC levels in the range of .05% (see Clayton, 1980; Gawron, and Ranney, 1988; Moskowitz, 1971, 1974 for reviews). Speed maintenance, cornering stability, braking distance, and fine psychomotor control movements all seem to be degraded by alcohol (see Gawron, and Ranney, 1988). It is these "real world" examples of alcohol's effects that reinforce the view that alcohol not only influences job-related behavior, but also the components or subtasks that make up more complex job performance. These simple subtasks not only serve as laboratory tasks for experimentally testing the influence of alcohol and other risk factors, but also serve as a task assortment from which to draw candidate RTP measures. It is the effects of alcohol on these tasks that will be reviewed next, following a brief look at the psychophysiology of alcohol.

Psychophysiology of Alcohol

Ethanol or ethyl alcohol, that form of alcohol most often associated with common beverages, is a colorless liquid with low molecular weight and nearly infinite water solubility. As a result of these rather unique characteristics, alcohol is absorbed directly through the oral tissues in the mouth, as well as the lining of the stomach and the intestinal tract. Absorption is so rapid that unless taken in large amounts, very little alcohol passes the duodenum (see Ritchie, 1985; Forney and Hughes, 1968).

Once consumed and absorbed, alcohol is transported rapidly throughout the body and to the brain. While alcohol has been shown to have numerous effects on biological processes including heat regulation, the circulatory process, gastric secretion, and diuresis, no effects are as profound as those seen in the central nervous system. Alcohol acts rapidly and effectively as a CNS depressant freeing numerous areas of the brain from inhibitory control. It is this process that gives rise to the greater disinhibition of action that is often the basis for confusing alcohol with stimulant drugs. In fact, alcohol acts to depress both excitatory and inhibitory postsynaptic potentials, as well as influencing numerous other neurophysiological functions (see Klemm, 1979).

It has been demonstrated that alcohol exerts its effects first on parts of the brain invested with integrative function such as the numerous cortical structures and the reticular activating system (Himwich and Callison, 1972). The result is that, with increasing consumption of ethanol, normally organized mental processes become disorganized, and motor processes become disrupted (see Ritchie, 1985). It is no doubt that these processes give rise to the changes in performance that are observed for cognitive and behavioral tasks after the ingestion of alcohol.

Chronic versus Acute Effects of Alcohol

Before reviewing some of the literature on alcohol effects on task performance, it is important to distinguish between chronic, long-term effects that are more often associated with alcoholism, and the acute, short-term effects of alcohol. Wechsler raised this important distinction between acute and chronic effects of alcohol as early as 1940 (see Wechsler, 1958).

Aside from the development of organic brain syndromes in cases of extended alcohol consumption, the effects of *chronic* alcohol consumption on gross intellectual function, memory, and learning ability appear to be moderate to mild, especially in contrast with the effects of acute episodes (Kleinknecht and Goldstein, 1972; Parsons and Leber, 1981). For example, full-scale IQ does not appear to be dramatically affected as a result of chronic alcoholism unless there is gross organic brain syndrome (Wechsler, 1958; Halpern, 1946; Murphy, 1953; Peters, 1956; Plumeau, Machover, and Puzzo, 1960). An examination of subscale performance reveals more noticeable differences however. Alcoholics perform less well on many of the *performance* subtests of the WAIS, as compared to nonalcoholics (Wechsler, 1958; see also Parsons and Leber, 1981). The locus of chronic alcohol influences seems to be centered on problem solving functions. As compared to nonalcoholics, alcoholics have been shown to perform more poorly on measures of abstract reasoning ability, perceptual-spatial-motor ability, and other measures of problem solving ability (Fitzhugh, Fitzhugh, and Reitan, 1965; Goldstein and Shelly, 1971; Parson and Leber, 1981; Williams and Skinner, 1990), especially those involving conceptual shifting (Tarter and Parsons, 1971). There is also some evidence that memory processes, perhaps during the initial acquisition phase, are also negatively influenced by chronic alcohol consumption (Nixon, Kujawski, Parsons, and Yohman, 1987).

The literature on *chronic* alcohol effects suggests fairly clear evidence of alcohol influences on one's ability to perform task functions. While helpful in confirming the potential of alcohol to disrupt task performance, these chronic effects are typically manifested in more enduring behavioral changes that would not be easily detected by RTP assessment unless reflected in a gradual shift in the individual's baseline. Assessment with RTP tests is most sensitive to transient changes in performance due to *acute* drug effects. More important for supporting the concept of RTP are investigations of the acute or short-term effects of alcohol on specific types of cognitive and psychomotor task performance.

Acute Effects of Alcohol on Task Performance

Memory. A very large number of studies have explored short-term alcohol effects on memory performance. This area of alcohol research is important to the concept of RTP because so many performance tasks depend on both short-term and long-term storage and retrieval processes (e.g., Sternberg, linguistic processing, grammatical reasoning, math processing, etc.).

Both short-term recall and recognition processes that underlie such tasks as the Sternberg, math processing, linguistic processing, grammatical reasoning, and spatial processing tasks appear to be degraded by alcohol. For example, both speed and accuracy of word recognition has been shown to be degraded by alcohol (Maylor and Rabbitt, 1987a; Maylor, Rabbitt, and Kingstone, 1987; Maylor, Rabbit, James, and Kerr, 1990), as well as recognition for pictures (Ryback, Weinert, and Fozard, 1970). Free recall of text and spatial information also seems to be degraded both in speed and accuracy by alcohol consumption (Jubis, 1990; Maylor, Rabbitt, and Kingstone, 1988; Maylor, Rabbit, James, and Kerr, 1990). Jones and Jones (1977) also demonstrated that alcohol appears to disrupt the storage process of early components in the memory set (i. e., primacy effects) as opposed to latter components (i. e., recency effects). Similar results of the disrupted memory processes of early components were reported by Hockey, MacLean, and Hamilton (1981). It should also be noted that a number of these investigations found negative alcohol effects on memory processes at BAC's in the range of .02 to .06% (Jubis, 1990; Ryback, Weinert, and Fozard, 1970).

These research results on alcohol and memory processes have been reviewed and interpreted in a more global context (see Maylor and Rabbitt, 1987a, discussion of Birnbaum and Parker, 1977). If a linear sequential processing model of memory is adopted with stages of encoding, storage, and retrieval, the most marked effects of alcohol are believed to be in the storage stage. Some explanations for the negative effects that alcohol has on memory have been offered, namely, that alcohol may decrease rehearsal ability, or that it may disrupt encoding (Craik, 1977), or increase forgetting (Wickelgren, 1975). It has also been suggested that alcohol might simply reduce motivation (Landauer, 1977). However, Forney and Hughes (1968, p.30) have suggested that alcohol might increase motivation and thus improve other types of performance.[1]

Despite the considerable evidence that these memory processes are disrupted by alcohol, the four-choice variant of the Sternberg task does not appear to show a sensitivity to alcohol effects (Stokes, Belger, Banich, and Taylor, 1991; Taylor, Dellinger, and Schilling, 1983). Carpenter and Ross (1965), however, have shown degraded performance as a result of alcohol on a continuous processing task, a task requiring the subject to store in memory information from previous trials for comparison to future trials. Also, Schlegel and Storm (1983) have reported degraded response accuracy on the manikin task, a test of spatial rotation ability, as a function of alcohol level.

One interesting result with respect to the recall memory literature is that a number of studies identified *facilitative* effects of alcohol on recall performance (Kalin, 1964; Lamberty, Beckwith, Petros, and Ross, 1990), recognition performance (Parker, Birnbaum, Weingartner, Hartley, Stillman, and Wyatt, 1980; Parker, Morihisa, Wyatt, Schwartz, Weingartner, and Stillman, 1981), and continuous processing (Carpenter and Ross, 1965). The alcohol doses in these studies ranged from .03 to 0.1%. One critical variable explaining these facilitative effects appears to be the point in the experimental protocol at which the alcohol is consumed. In most of the memory research cited previously, subjects were administered alcohol and brought to the target BAC *before* they were tested (i.e., pretrial). In the studies showing facilitative effects, the subjects were trained or exposed to the material to be learned prior to exposure to alcohol (i. e., post-trial). Why post-trial administered alcohol provides facilitative effects on memory is not clear, but both interference theory and consolidation theory explanations have been offered (see Lamberty, Beckwith, Petros, and Ross, 1990). Note, however, that facilitative effects have been anecdotally cited in studies without post-trial alcohol administration (see Collins, 1979). In this regard, one additional theory of alcohol's facilitative effects is intriguing. Goldberg (1969; see also Pohorecky, 1977) has suggested that alcohol at BAC's ranging from .02 to .03% acts as a central nervous system stimulant thereby facilitating performance through generalized arousal mechanisms. However, this arousal hypothesis has not been given universal support (Gustafson, 1987a).

Memory processes are most certainly affected by alcohol, While the research evidence is not entirely consistent, many of these studies suggest that storage processes are disrupted by alcohol. This apparently affects task behavior in the form of lengthened response times and decreased accuracy. These disruptive effects in basic component tasks may, in turn, combine to produce some of the observed differences in more complex behavior.

---

1 There is another alternative explanation for not only the disruption seen on memory tasks, but also on other tasks as well. Many of the performance decrements seen following the administration of alcohol could be due simply to the negative influence that alcohol has on the visual processes (see Forney & Hughes, 1968). Double vision, nystagmus, blurring of color vision, loss of depth perception, loss of acuity, and loss of the fusion reflex needed for binocular vision are all affected negatively by alcohol consumption (Aschan, Bergstedt, Goldberg, & Laurell, 1956; Bjerver & Goldberg, 1951; Wist, Hughes, & Forney, 1967).

Response Speed. Among the many measures used to assess human performance are response speed (i. e., reaction time or response time) measures. Response speed measures are popular because they are among the simplest measures to record, they require less training, and they provide a considerable amount of information regarding the efficiency with which a person is responding. Once learned and well-practiced, response speed tasks can be very sensitive indicators of not only basic psychomotor ability, but also cognitive processes as well. The two measures addressed in this section of the review are simple reaction time and choice reaction time.

There is considerable variation in the manner in which reaction time measures are implemented. Distinctions should be drawn between reaction time and response time, the latter usually being a more complex measure including both reaction time and movement time. A distinction is also usually made between *simple* reaction time or response time and choice response time. "Simple" response time (SRT) measures typically involve one circumscribed psychomotor response to one sensory stimulus. A common variation is the "serial" response measure in which there is a continuous presentation of stimuli, as opposed to discrete trials.

Choice response time (CRT) measures include two or more responses, among which the subject must choose, thereby requiring more decision and or information processing skills to be entered into the total response. For the purposes of this report, the research literature on response speed will be divided into the broad categories of SRT and CRT measures.

There have been numerous explorations of alcohol effects on SRT measures. Reviews of these investigations were conducted as early as 1940 (Jellinek and McFarland; see also Carpenter, 1962). These reviews concluded that SRT across a wide range of stimulus types is generally degraded by alcohol. However, the exact nature and degree of the degradation was difficult to assess given the wide variation in methodologies applied.

More recent studies of alcohol effects on SRT have added some clarification. These studies reaffirmed that, in general, SRT is negatively affected by alcohol consumption. But, an important variable appears to be the amount of alcohol consumed (see Dinges and Kribbs, 1989). Of over twenty studies conducted since the early 1960s, most found negative effects of alcohol on SRT at alcohol dose levels producing BAC's of approximately .07% or above (e.g., Collins, 1979; Smith, Sinha, and Williams, 1989; Sutton and Burns, 1971; Sutton and Kimm, 1970). The predominant finding is that alcohol at this level or above appears to affect SRT by slowing mean response time. Trial length may also play a role in understanding alcohol effects. It has been suggested that trial lengths less than five minutes may not be long enough to reveal alcohol effects (Dinges and Kribbs, 1989). This may be due to increased habituation of SRT under the influence of alcohol at higher levels (Gustafson, 1986a, 1986b).

Choice reaction time studies have provided fairly consistent evidence for the negative effects of alcohol on performance speed. This is not surprising because CRT tasks incorporate greater demands on decision and other information processes as compared to SRT tasks. In fact, based on a review of much of the earlier literature, Perrine (1976) concluded that CRT was more sensitive to the negative effects of alcohol than SRT. More recent research has confirmed these findings. Alcohol does seem to have a negative effect on CRT, even at moderate doses (Gustafson, 1987b; Hindmarch, Kerr, and Sherwood, 1991; Maylor and Rabbitt, 1987b; Maylor, Rabbitt, and Connolly, 1989; Maylor, Rabbitt, Sahgal, and Wright, 1987). It appears that this negative effect is oriented primarily toward slowing the processing of an event requiring a choice -- as opposed to slowing the rate at which one prepares for a choice response event (Maylor and Rabbitt, 1989).

Tracking. The influence of alcohol on tracking performance has been studied in numerous investigations, and the results of these investigations are reasonably consistent. In general, tracking performance is markedly degraded both during resting conditions (Collins, 1979; Collins and Chiles, 1979; Collins, Schroeder, Gilson, and Guedry, 1971; Chiles and Jennings, 1969; Dott and McKelvy, 1977; Gilson, Schroeder, Collins, and Guedry, 1972; Klein and Jex, 1975; Schroeder, Gilson, Guedry, and Collins, 1973; Stokes, Belger, Banich, and Taylor, 1991) and during acceleration (Collins and Chiles, 1979). There is even evidence that negative alcohol effects on tracking performance may occur at BAC's of .03 to .05% (Gilson, Schroeder, Collins, and Guedry, 1972). Evidence for negative alcohol effects on related behavior, such as maze tracing, has also been presented (Stokes, Belger, Banich, and Taylor, 1991).

Vigilance. Vigilance performance forms the basis for many types of jobs requiring sustained monitoring for the occurrence of critical, yet low frequency, events (e. g., radar operations or control room operations). Evidence for vigilance degradation due to alcohol comes from several investigations based on serial response tasks (e. g., Gustafson, 1986c), meter detection (Chiles and Jennings, 1969), and specially designed vigilance tasks (e. g., Wilkinson Auditory Vigilance Task; Horne and Gibbons, 1991). Alcohol appears to lengthen the time needed to detect critical events, although the dynamics of this process are not completely clear. It has been reported that alcohol seems to interact with time on task. That is, the longer one remains on task under the influence of alcohol, the greater alcohol negatively affects detection time (Gustafson, 1986c). Some degradation of vigilance ability, especially for responses that would normally be among someone's slowest responses, seems to occur even at BAC levels of about .06% (Gustafson, 1986d). Again, there is some evidence that short time periods (under five minutes) may not be affected by alcohol as much as longer time periods (Dinges and Kribbs, 1989; Gustafson, 1986c). And, time-of-day may also be an important variable in regulating the effects of alcohol on vigilance performance. Horne and Gibbins (1991) have reported that alcohol causes greater negative effects on vigilance performance in the early afternoon than in the evening.

Complex Task Performance. There is limited, yet reasonably consistent evidence that alcohol has negative effects on complex task performance. For the purposes of this report, complex tasks include those tasks where more than one task is performed simultaneously. For example, Miles, Porter and Jones (1986) administered simultaneously a tracking task and a version of the Bakan vigilance task. The results of this study suggested that alcohol had marginally significant negative effects on tracking and on vigilance performance. The authors suggested that the negative effects were stronger for the vigilance task, as opposed to the tracking task -- consistent with many but not all reviews of the alcohol performance literature (see Levine, Kramer, and Levine, 1975).

Other investigations of alcohol effects on complex task performance have utilized sophisticated, multitask batteries. In these investigations, the subjects perform a complex task that might include many more than two tasks. For example, a subject might be called upon to perform simultaneously a monitoring, a pattern discrimination, a mental arithmetic, and a tracking task (c.f., Collins, Mertens, and Higgins, 1987). Results of studies utilizing multitask batteries reveal that alcohol has negative effects on complex task performance (Collins, 1980; Collins and Chiles, 1979; Collins and Mertens, 1988; Collins, Mertens, and Higgins, 1987). These studies also explored the interaction with simulated altitude and found no compounding effect of altitude with alcohol. That is, both altitude and alcohol had similar negative effects on complex task performance, but there was no synergistic interactive effect. In general, the tasks that have consistently proven to be most sensitive to low blood alcohol levels involve multi-tasking or divided attention..

The research on alcohol effects on performance seem to support the view that alcohol has its most serious consequences for those tasks that require cognitive or information processing ability. The more these processes are demanded or the more sophisticated the processes are, the more alcohol seems to disrupt functioning. Regardless of the type of task, when BAC reaches 0.10%, most task performance is affected regardless of its nature. It also appears that the longer one maintains task performance under the influence of alcohol the greater the likelihood that negative effects will develop. There is also some suggestion that short periods of task performance, even for cognitive/information processing tasks, may not be sufficient to reveal alcohol effects. Alcohol seems to have negative effects on many of the tasks commonly represented in modern task batteries, i.e., Sternberg, digit span, linguistic processing, spatial processing, tracking, and various response time and vigilance tasks. Whether these tasks are presented individually or in combination, alcohol consumption appears to affect them negatively.

OTHER DRUGS

The effects of a number of other drugs on human performance have been reported. The literature on these drugs is not nearly as extensive as the literature on alcohol. However, a brief overview of each drug follows.

Antihistamines. Antihistamines are taken in large quantities in the treatment of common allergies and in cold medications. Their main side-effect is drowsiness. In addition, they have been shown to have serious negative influences on a broad range of human performance tasks including reaction time, tracking, continuous memory, visual search, digit symbol, divided attention, and vigilance (see Eddy, Dalrymple, and Schiflett, 1992; and Nesthus, Schiflett, Eddy, and Whitmore, 1991 for brief reviews).

Caffeine. Many contradictory findings regarding the effect of caffeine on task performance have been reported. Based on a number of reviews (Gilbert, 1976; Gilliland and Bullock, 1983; Weiss and Laties, 1962), several conclusions can be drawn. First, the effect of caffeine on reaction time (RT) is complex. Very low doses do not seem to have much of an effect. However, moderate doses 200-300mg) appear to facilitate RT, but then may degrade performance when subjects are tested 24 hours later. Caffeine abstainers probably benefit most from low-dose facilitative effects. Other measures such as mathematical processing, coding, complex verbal performance, and skilled psychomotor tasks are facilitated by low doses of caffeine, especially if boredom or fatigue is a factor.

Cocaine. The literature on cocaine is meager and contradictory. There have been some reports that cocaine has a facilitative effect on vigilance, psychomotor performance, and memory retrieval. However, it appears that such effects exist only for fairly simple tasks and that as tasks become more complex, this effect is negligible (see Byck, 1987; Ellinwood and Nikaido, 1987). One additional effect that cocaine appears to have is the ability to overcome fatigue. Cocaine does not appear to enhance performance so much as it seems to overcome the degrading effects of fatigue or sleep deprivation (Fischman and Schuster, 1980).

FATIGUE AND SLEEP LOSS

The effects of fatigue and sleep loss (including such areas as circadian shift-work effects) have been studied extensively. A complete review of each area is beyond the scope of this review. However, there are some commonalities across these areas with regard to their effects on performance. Fatigue and sleep loss appear to increase the onset and frequency of decrements in vigilance tasks, memory tasks, logical reasoning tasks, mathematical reasoning tasks, and complex verbal processing and decision making tasks (see Krueger, 1989). Many of the negative effects are probably the result of decreased efficiency in detecting visual and auditory signals (Stroh, 1971; Mackie, 1977). It also appears that the severity of these decrements is worse during prolonged or boring tasks. However, others have noted that motivation can play an important role in overcoming these decrements (Wilkinson, 1964). Specifically, adequate motivation can allow subjects to perform well during extended periods of fatigue or sleep loss.

SUMMARY

This review has attempted to present examples of the types of literature that link risk factors to patterns of responding on specific types of human performance tasks. It is fairly clear that an abundant amount of literature exists that shows risk factors to have either positive or negative effects on specific forms of cognitive and psychomotor performance. While some of this literature has been used to support the claims of validity for specific RTP measures, it might be said that the best use of such literature is to aid in the process of initially identifying *candidate* RTP measures that should then each be independently validated for the specific purpose to which they will be applied.

REFERENCES

Aschan, G., Bergstedt, M., Goldberg, L., and Laurell, L. (1956). Positional nystagmus in man during and after alcohol intoxication. *Quarterly Journal of Studies on Alcohol, 17,* 381-405.

Billings, C. E., Wick, R. L., Gerke, R. J., and Chase, R. C. (1973). Effects of ethyl alcohol on pilot performance. *Aerospace Medicine, 44,* 379-382.

Birnbaum, I. M., and Parker, E. S. (1977). Acute effects of alcohol on storage and retrieval. In I. M. Birnbaum and E. S. Parker (Eds.), *Alcohol and human memory.* New Jersey: Lawrence Erlbaum.

Bjerver, K., and Goldberg, L. (1951). Alcohol and road traffic. In *Proceedings of the First International Conference on Alcohol and Road Traffic,* Stockholm: Kugelbergs Boktryeri.

Byck, R. (1987). The effects of cocaine on complex performance in humans. *Alcohol, Drugs, and Driving,* 3, 9-12.

Carpenter, J. A. (1962). Effects of alcohol on some psychological processes. A critical review with special reference to automobile driving skill. *Journal of Studies on Alcohol, 39,* 274-314.

Carpenter, J. A., and Ross, B. M. (1965). Effect of alcohol on short-term memory. *Quarterly Journal of Studies on Alcohol, 26,* 561-579.

Chiles, D., and Jennings, A. E. (1969). *Effects of alcohol on complex performance* (Technical Report-AM-69-14). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Clayton, A. B. (1980). Effects of alcohol on driving skills. In M. Sandler (Ed.), *Psychopharmacology of alcohol.* New York: Raven.

Collins, W. E. (1979). *Performance effects of alcohol intoxication and hangover at ground level and at simulated altitude* (Technical Report-AM-79-26). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Collins, W. E. (1980). Performance effects of alcohol intoxication and hangover at ground level and at simulated altitude. *Aviation, Space, and Environmental Medicine, 51,* 327-335.

Collins, W. E., and Chiles, W. D. (1979). *Laboratory performance during acute intoxication and hangover* (Technical Report-AM-79-7). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Collins, W. E., and Mertens, H. W. (1988). Age, alcohol, and simulated altitude: Effects on performance and breathalyzer scores. *Aviation, Space, and Environmental Medicine, 59,* 1026-1033.

Collins, W. E., Mertens, H. W., and Higgins, E. A. (1987). Some effects of alcohol and simulated altitude on complex performance scores and breathalyzer readings. *Aviation, Space, and Environmental Medicine, 58,* 328-332.

Collins, W. E., Schroeder, D. J., Gilson, R. D., and Guedry, F. E. (1971). Effects of alcohol ingestion on tracking performance during angular acceleration. *Journal of Applied Psychology, 55,* 559-563.

Craik, F. M. I. (1977). Similarities between the effects of age and alcoholic intoxication on memory performance, construed within 'levels of processing' framework. In I. M. Birnbaum and E. S. Parker (Eds.), *Alcohol and human memory.* New Jersey: Lawrence Erlbaum.

DHHS -- Department of Health and Human Services (1990). *Seventh special report to the U.S. Congress on alcohol and health from the Secretary of Health and Human Services* (DHHS ADM-281-88-0002). Rockville, MD: DHHS.

Dinges, D. F., and Kribbs, N. B. (1989). Comparison of the effects of alcohol and sleepiness on simple reaction time performance: Enhanced habituation as a common process. *Alcohol, Drugs, and Driving, 5,* 341-350.

Dott, A. B., and McKelvy, R. K. (1977). Influence of ethyl alcohol in moderate levels on visual stimulus tracking. *Human Factors, 19,* 191-199.

Eddy, D. R., Dalrymple, M. A., and Schiflett, S. G. (1992). *Comparative effects of antihistamines on aircrew mission effectiveness under sustained operations* (Technical Report AL-TR-1992-0018). Brooks Air Force Base, TX: Armstrong Laboratory.

Ellinwood, E. H., and Nikaido, A. M. (1987). Stimulant induced impairment: A perspective across dose and duration of use. *Alcohol, Drugs, and Driving, 3,* 19-24.

Fischman, M. W., and Schuster, C. R. (1980). Cocaine effects in sleep-deprived humans. *Psychopharmacology, 72,* 1-8.

Fitzhugh, L. C., Fitzhugh, K. B., and Reitan, R. M. (1965). Adaptive abilities and intellectual functioning of hospitalized alcoholics: Further considerations. *Quarterly Journal of Studies on Alcohol, 26,* 402-411.

Forney, R. B., and Hughes, F. W. (1968). *Combined effects of alcohol and other drugs.* Springfield, IL: Charles C. Thomas.

Gawron, V. J., and Ranney, T. A. (1988). The effects of alcohol dosing on driving performance on a closed course and in a driving simulator. *Ergonomics, 31*, 1219-1244.

Gibbons, H. L. (1988). Alcohol, aviation, and safety revisited: a historical review and suggestion. *Aviation, Space and Environmental Medicine, 59*, 657-660.

Gilbert, R. M. (1976). Caffeine as a drug of abuse. In R. G. Gibbin, Y. Isreal, H. Kalant, R. E. Popham, W Schmidt, and R. G. Smart (Eds.), *Research advances in alcohol and drug problems, Vol 3*, New York: Wiley.

Gilliland, K., and Bullock, W. (1983). Caffeine: A potential drug of abuse. *Advances in Alcohol and Substance Abuse, 3*, 53-73.

Gilson, R. D., Schroeder, D. J., Collins, W. E., and Guedry, F. E. (1972). Effects of different alcohol dosages and display illumination on tracking performance during vestibular stimulation. *Aerospace Medicine, 43*, 656-660.

Goldberg, L. (1969). Physiological research on alcohol, 1960-1970. In M. Keller and T. Coffey (Eds.), *Proceedings of the 28th International Congress on Alcohol and Alcoholism, Vol 2*. Highland, NJ: Millhouse Press.

Goldstein, G., and Shelly, C. H. (1971). Field dependence and cognitive, perceptual and motor skills in alcoholics. *Quarterly Journal of Studies on Alcohol, 32*, 29-40.

Gustafson, R. (1986a). Alcohol and vigilance performance: Effect of small doses of alcohol on simple visual reaction time. *Perceptual and Motor Skills, 62*, 99-102.

Gustafson, R. (1986b). Effect of small doses of alcohol and signal intensity on simple auditory reaction time in a monotonous test situation. *Perceptual and Motor Skills, 62*, 951-955.

Gustafson, R. (1986c). Effect of moderate doses of alcohol on simple auditory reaction time in a vigilance setting. *Perceptual and Motor Skills, 62*, 683-690.

Gustafson, R. (1986d). Alcohol and vigilance performance: Effect of small doses of alcohol on simple auditory reaction time. *Perceptual and Motor Skills, 63*, 99-102.

Gustafson, R. (1987a). Effect of alcohol and setting on three mood dimensions. *Psychological Reports, 60*, 503-507.

Gustafson, R. (1987b). Reaction time as a function of alcohol and selective attention. *Journal of Social Behavior and Personality, 2*, 515-522.

Halpern, F. (1946). Studies of compulsive drinkers: Psychological test results. *Quarterly Journal of Studies on Alcohol, 6,* 468-479.

Harper, C. R., and Albers, W. R. (1964). Alcohol and general aviation accidents. *Aerospace Medicine, 35,* 462-464.

Henry, P. H., Flueck, J. A., and Sanford, J. F. (1974). Assessment of performance in a Link GAT-1 flight simulator at three alcohol dose levels. *Aerospace Medicine, 45,* 33-44.

Higgins, E. A., Vaughan, J. A., and Funkhouser, G. E. (1970). *Blood alcohol concentration of alcoholic beverage dosages and altitudes* (Technical Report-AM-70-5). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Himwich, H. E., and Callison, D. A. (1972). The effects of alcohol on evoked potentials of various parts of the central nervous system of the cat. In B. Kissin and H Begleiter (Eds.), *The biology of alcoholism: Vol 2 Physiology and behavior* . New York: Plenum.

Hindmarch, I., Kerr, J. S., and Sherwood, N. (1991). The effect of alcohol and other drugs on psychomotor performance and cognitive function. *Alcohol and Alcoholism, 26,* 71-79.

Hockey, R., MacLean, A., and Hamilton, P. (1981). State changes and the temporal patterning of component resources. In J. Long and A. Baddeley (Eds.), *Attention and performance: IX.* New Jersey: Lawrence Erlbaum.

Horne, J. A., and Gibbons, H. (1991). Effects on vigilance performance and sleepiness of alcohol given in the early afternoon ('post lunch') vs. early evening. *Ergonomics, 4,* 67-77.

Jellinek, E. M., and McFarland, R. A. (1940). Analysis of psychological experiments on effects of alcohol. *Journal of Studies on Alcohol, 1,* 272-371.

Jones, B. M., and Jones, M. K. (1977). Alcohol and memory impairment in male and female social drinkers. In I. M. Birnbaum and E. S. Parker (Eds.), *Alcohol and human memory.* New Jersey: Lawrence Erlbaum.

Jubis, R. M. T. (1990). Effects of alcohol and white noise on recall of relevant and irrelevant task components. *Perceptual and Motor Skills, 71,* 691-702.

Kalin, R. (1964). Effects of alcohol on memory. *Journal of Abnormal and Social Psychology, 69,* 635-641.

Klein, R. H., and Jex, H. R. (1975). Effects of alcohol on a critical tracking task. *Journal of Studies on Alcohol, 36,* 11-20.

Kleinknecht, R. A., and Goldstein, S. G. (1972). Neuropsychological deficits associated with alcoholism. *Quarterly Journal of Studies on Alcohol, 33*, 999-1019.

Klemm, W. R., (1979). Effects of ethanol on nerve impulse activity. In E. Majchrowicz and E. P. Noble (Eds.), *Biochemistry and pharmacology of ethanol, Vol 2.* New York: Plenum.

Kreuger, G. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress, 3*, 129-141.

Lamberty, G. J., Beckwith, B. E, Petros, T. V., and Ross, A. R. (1990). Posttrial treatment with ethanol enhances recall of prose narratives. *Physiology and Behavior, 48*, 653-658.

Landauer, T. K. (1977). Remarks on the detection and analysis of memory deficits. In I. M. Birnbaum and E. S. Parker (Eds.), *Alcohol and human memory.* New Jersey: Lawrence Erlbaum.

Levine, J. M., Kramer, G. G., and Levine, E. N. (1975). Effects of alcohol on human performance: An integration of research findings based on an abilities classification. *Journal of Applied Psychology, 60*, 285-293.

Mackie, R. R. (1977). *Vigilance: Theory, operational performance and physiological correlates.* New York: Plenum.

Maylor, E. A., and Rabbitt, P. M. A. (1987a). Effect of alcohol on rate of forgetting. *Psychopharmacology, 91*, 230-235.

Maylor, E. A., and Rabbitt, P. M. A. (1987b). Effects of alcohol and practice on choice reaction time. *Perception and Psychophysics, 42*, 465-475.

Maylor, E. A., and Rabbitt, P. M. A. (1989). Relationship between rate of preparation for, and processing of, an event requiring a choice response. *Quarterly Journal of Experimental Psychology, 41*, 47-62.

Maylor, E. A., Rabbitt, P. M. A., and Connolly, S. A. V. (1989). Rate of processing and judgement of response speed: Comparing the effects of alcohol and practice. *Perception and Psychophysics, 45*, 431-438.

Maylor, E. A., Rabbitt, P. M. A., and Kingstone, A. F. (1987). Effect of alcohol on word categorization and recognition memory. *British Journal of Psychology, 78*, 233-239.

Maylor, E. A., Rabbitt, P. M. A., and Kingstone, A. F. (1988). Effect of alcohol on lexical access. *Psychopharmacology, 95*, 119-123.

Maylor, E. A., Rabbitt, P. M. A., James, G. H., and Kerr, S. A. (1990). Comparing the effects of alcohol and intelligence on text recall and recognition. *British Journal of Psychology, 81*, 299-313.

Maylor, E. A., Rabbitt, P. M. A., Sahgal, A., and Wright, C. (1987). Effects of alcohol on speed and accuracy in choice reaction time and visual search. *Acta Psychologica, 65*, 147-163.

Mello, N. K., and Mendelson, J. H. (1978). Alcohol and human behavior. In L. L. Iversen and S. H. Snyder (Eds.), *Handbook of psychopharmacology, Vol. XII.* New York: Plenum.

Modell, J. G., and Mountz, J. M. (1990). Drinking and flying -- The problem of alcohol use by pilots. *The New England Journal of Medicine, 323*, 455-461

Moskowitz, H. (1971). *The effects of alcohol on performance in a driving simulator of alcoholics and social drinkers* (Report ENG-7205). Institute of Transportation and Traffic Engineering, UCLA.

Moskowitz, H. (1974). Validity of driving simulator studies for predicting drug effects in real driving situations. In *Proceedings of the 6th International Conference on Alcohol, Drugs and Traffic Safety*, 295-303.

Moskowitz, H., and Robinson, C. (1988). *Effects of low doses of alcohol on driving-related skills: A review of the evidence.* Washington, DC:U.S. Department of transportation.

Murphy, M. M. (1953). Social class differences in intellectual characteristics of alcoholics. *Quarterly Journal of Studies on Alcohol, 14*, 192-196.

Nathan, P. E. (1990). Residual effects of alcohol. *National Institute on Drug Abuse Research Monograph Series, Monograph 101*, 112-123.

Nesthus, T. E., Schiflett, S. G., Eddy, D. R., and Whitmore, J. N. (1991). *Comparative effects of antihistamines on aircrew performance of simple and complex tasks under sustained operations* (Technical Report AL-TR-1991-0104). Brooks Air Force Base, TX: Armstrong Laboratory.

Nixon, S. J., Kujawski, A., Parsons, O. A., and Yohman, J. R. (1987). Semantic (verbal) and figural memory impairment in alcoholics. *Journal of Clinical and Experimental Neuropsychology, 9*, 311-322.

Parker, E. S., Birnbaum, I. M., Weingartner, H., Hartley, J. T., Stillman, R. C., and Wyatt, R. J. (1980). Retrograde enhancement of human memory with alcohol. *Psychopharmacology, 69*, 219-222.

Parker, E. S., Morihisa, J. M., Wyatt, R. J., Schwartz, B. L., Weingartner, H., and Stillman, R. C. (1981). The alcohol facilitation effect on memory: A dose-response study. *Psychopharmacology, 74,* 88-92.

Parsons, O. A., and Leber, W. R. (1981). The relationship between cognitive dysfunction and brain damage in alcoholics: Causal, interactive, or epiphenomenal? *Alcoholism: Clinical and Experimental Research, 5,* 326-343.

Perrine, M. W. (1976). Alcohol and highway crashes: Closing the gap between epidemiology and experimentation. *Modern Problems in Pharmacopsychiatry, 11,* 22-41.

Peters, G. A. (1956). Emotional and intellectual concomitants of advanced chronic alcoholism. *Journal of Consulting Psychology, 20,* 390.

Plumeau, F., Machover, S., and Puzzo, F. (1960). Wechsler-Bellevue performances of remitted and unremitted alcoholics, and their normal controls. *Journal of Consulting Psychology, 24,* 237-248.

Pohorecky, L. A. (1977). Biphasic action of ethanol. *Biobehavioral Review, 1,* 231-240.

Ritchie, J. M. (1985). The aliphatic alcohols. In A. G. Gilman, L. S., Goodman, T. W., Rall, and F. Murad (Eds.), *The pharmacological basis of therapeutics.* New York: Macmillan.

Ross, L. E., and Mundt, J. C. (1988). Multiattribute modeling analysis of the effects of a low blood alcohol level on pilot performance. *Human Factors, 30,* 293-304.

Ross, L. E., Yeazel, L. M., and Chau, A. W. (1992). Pilot performance with blood alcohol concentrations below 0.04%. *Aviation, Space, and Environmental Medicine, 63,* 951-956.

Ryback, R. S. (1970). Effects of alcohol on memory and its implications for flying safety. *Aerospace Medicine, 41,* 1193-1195.

Ryback, R. S., Weinert, J., and Fozard, J. L. (1970). Disruption of short-term memory in man following consumption of ethanol. *Psychonomic Science, 20,* 353-354.

Schlegel, R. E., and Storm, W. F. (1983). Speed-accuracy tradeoffs in spatial orientation information processing. In *Proceedings of the 27th Annual Meeting of the Human Factors Society,* 359-363.

Schroeder, D. J., Gilson, R. D., Guedry, F. E., and Collins, W. E. (1973). Effects of alcohol on nystagmus and tracking performance during laboratory angular accelerations about the Y and Z axes. *Aerospace Medicine, 44,* 477-487.

Smith, L. T., Sinha, R., and Williams, H. L. (1989). The interaction of alcohol and sleep deprivation in two reaction time tasks. *Alcohol, Drugs, and Driving, 5,* 341-350.

Stokes, A. F., Belger, A., Banich, M. T., and Taylor, H. (1991). Effects of acute aspartame and acute alcohol ingestion upon the cognitive performance of pilots. *Aviation, Space and Environmental Medicine, 62,* 648-653.

Stroh, C. M. (1971). *Vigilance: The problem of sustained attention.* Oxford: Pergamon.

Sutton, D., and Burns, J. (1971). Alcohol dose effects on feedback-maintained simple reaction time. *Journal of Psychology, 78,* 151-159.

Sutton, D., and Kimm, J. (1970). Alcohol effects on human motor unit reaction time. *Physiology and Behavior, 5,* 889-892.

Tang, P. C., and Rosenstein, R. (1967). Influence of alcohol and dramamine, alone and in combination, on psychomotor performance. *Aerospace Medicine, 38,* 818-821.

Tarter, R. E., and Parsons, O. (1971). An analysis of the cognitive deficits in chronic alcoholics. *Journal of Abnormal Psychology, 77,* 71-75.

Taylor, H. L., Dellinger, J. A., and Schilling, R. L. (1983). Pilot performance measurement methodology for determining the effects of alcohol and other toxic substances. In *Proceedings of the 27th Annual Meeting of the Human Factors Society,* 334-338.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence (4th ed.)* Baltimore: Williams and Wilkins.

Weiss, B., and Laties, V. (1962). Enhancement of human performance by caffeine and amphetamines. *Pharmacological Review, 14,* 1-36.

Wickelgren, W. A. (1975). Alcoholic intoxication and memory storage dynamics. *Memory and Cognition, 3,* 385-389.

Wilkinson, R. T. (1964). Effect of up to 60 hours of sleep derivation on different types of work. *Ergonomics, 7,* 175-186.

Williams, C. M., and Skinner, A. E. G. (1990). The cognitive effects of alcohol abuse: A controlled study. *British Journal of Addiction, 85,* 911-917.

Wist, E. R., Hughes, F. W., and Forney, R. B. (1967). Effect of low blood alcohol level on stereoscopic acuity and fixation disparity. *Perceptual and Motor Skills, 24,* 83-87.

☆U.S. GOVERNMENT PRINTING OFFICE: 1993 - 761-068/80006