DOCUMENT RESUME

ED 250 320 TM 840 444

AUTHOR Hunt, Earl; Pellegrino, James
TITLE Using Interactive Computing to Expand Intelligence Testing: A Critique and Prospectus. Report No. 84-2.
INSTITUTION California Univ., Santa Barbara. Dept. of Psychology.; Washington Univ., Seattle. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE May 84
CONTRACT N00014-80-C-0631; N00014-81-C-0616
NOTE 85p.
PUB TYPE Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Attention; *Cognitive Tests; *Computer Assisted Testing; *Futures (of Society); Learning Processes; Logical Thinking; Memory; *Microcomputers; Psychomotor Skills; Spatial Ability; Verbal Ability

ABSTRACT

If microcomputers are used as automated testing stations, for use in psychometric assessment, there are economic advantages. Discussion follows, however, on whether it is possible to improve the quality of cognitive assessment by extending the range of cognitive abilities to be assessed. Two types of extension are considered: modifying and expanding testing procedures for psychological functions that are components of conventional tests, and the extension of testing to psychological functions not generally assessed by conventional intelligence or aptitude tests. Computerized presentations will make relatively little difference in the current ways of testing verbal comprehension. Computer controlled testing could well extend the ways in which spatial-visual reasoning and memory are evaluated. The impact of testing on the evaluation of reasoning is unclear. Computer-controlled item presentation makes it possible to conceive of tests of learning and attention. The psychological nature of the abilities being assessed raises problems in assessment that are not addressed by the fact of computer-control. Some research questions are identified that ought to be explored before testing is extended into these fields. Computer-controlled evaluation could be extended to the assessment of criterion performance, either in the normal working situation or in a simulation of the workplace. While evaluation of this sort does raise social questions, it clearly presents an opportunity to obtain validation data for psychological assessment studies. (Author/BW)

TM

ED250320

USING INTERACTIVE COMPUTING TO EXPAND INTELLIGENCE TESTING. A CRITIQUE AND PROSPECTUS.

EARL HUNT

AND

JAMES PELLEGRINO

MAY 1984

This research was sponsored by:

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research
Under Contract No. N00014-80-C-0631
and
Contract No. N00014-81-C-0616
Contract Authority Numbers NR 150-457
and NR 150-468.

TM 840 444

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 84-2	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): USING INTERACTIVE COMPUTING TO EXPAND INTELLIGENCE TESTING. A CRITIQUE AND PROSPECTUS (1)	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Earl Hunt and James Pellegrino	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0631
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, University of Washington, Seattle, WA 98195; Department of Psychology, University of California at Santa Barbara. 93106	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (code 442PT), Arlington, VA 22217	12. REPORT DATE: May, 1984
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for publ[illegible] release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Intelligence, computers, inter-active computing, psychometrics, verbal ability, spatial ability, attention, personnel selection, reasoning, intelligence tests.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Microcomputers can serve as automated testing stations for use in psychometric assessment. There are economic advantages in conducting aptitutde and intelligence testing with such stations. Is it possible to improve the quality of cognitive assessment by extending the range of cognitive abilities to be assessed? Two types of extension are considered; modifiying and expanding testing procedures for psychological functions that are components of conventional tests, and the extension of testing to psychological functions not generally assessed by conventional intelligence or aptitude tests. Computer-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) ized presentations will make relatively little difference in our ways of testing verbal comprehension. Computer controlled testing could well extend the ways in which we evaluate spatial-visual functioning and memory. The impact of testing on the evaluation of reasoning is unclear. Computer-controlled item presentation makes it possible to conceive of tests of learning and attention, neither of which are evaluated in most psychometric programs today. The psychological nature of the abilities being assessed raises problems in assessment that are irrelevant to the use of computers. Some research questions are identified that ought to be explored before testing is extended into these fields. Computer-controlled evaluation could be extended to the assessment of criterion performance, either in the normal working situation or in a simulation of the workplace. While evaluation of this sort does raise social questions, it clearly presents an opportunity to obtain validation data for psychological assessment studies.

Using Interactive Computing to Expand Intelligence Testing. A Critique and Prospectus(1)

Earl Hunt

The Universit of Washington

ai.

James Pellegrino

The University of California at Santa Barbara

Running Head: INTELLIGENCE TESTING

Introduction

We all know that we live in the computer era. At the same time an eminent science writer has called our culture the psychological society (Gross, 1978). Inevitably, the two have met. At the 1983 American Psychological Society meeting many vendors offered demonstrations of how various psychological functions could be evaluated on a computer. The Department of Defense is in the process of moving to computer controlled testing in recruit evaluation, one of the largest personnel selection programs in the world (Green et al., 1982). On the other hand, misgivings about the rush to computerize have been voiced. An example of such a misgiving is Matarazzo's (1983) concern that the introduction of computers could change both the nature of testing and the interpretation of test scores. His remarks serve as a useful reminder that personnel evaluation is at its heart an exercise in psychology, not computer science. Applying computer technology is a means, not an end.

In this paper we shall discuss some of the potentials and problems involved in computerized testing. There are two issues involved. One deals with the advantages or disadvantages of using a computer to administer "a test", regardless of the test's content. Most of the literature in the field has addressed this issue. The advantage seems to lie with the computer, for reasons that we describe immediately below. (We shall express a few misgivings). A far more difficult issue is whether or not the content of mental tests can be,

Introduction

We all know that we live in the computer era. At the same time an eminent science writer has called our culture the psychological society (Gross, 1978). Inevitably, the two have met. At the 1983 American Psychological Society meeting many vendors offered demonstrations of how various psychological functions could be evaluated on a computer. The Department of Defense is in the process of moving to computer controlled testing in recruit evaluation, one of the largest personnel selection programs in the world (Green et al., 1982). On the other hand, misgivings about the rush to computerize have been voiced. An example of such a misgiving is Matarazzo's (1983) concern that the introduction of computers could change both the nature of testing and the interpretation of test scores. His remarks serve as a useful reminder that personnel evaluation is at its heart an exercise in psychology, not computer science. Applying computer technology is a means, not an end.

In this paper we shall discuss some of the potentials and problems involved in computerized testing. There are two issues involved. One deals with the advantages or disadvantages of using a computer to administer "a test", regardless of the test's content. Most of the literature in the field has addressed this issue. The advantage seems to lie with the computer, for reasons that we describe immediately below. (We shall express a few misgivings). A far more difficult issue is whether or not the content of mental tests can be, or should be, changed in computerized testing. The second issue will occupy most of our discussion.

Tests of mental ability can be classified as individually administered or group administered tests. Individually administered tests are generally regarded as more valid, for two reasons. The examiner, functioning as a stimulus presentation device, can produce a variety of different auditory, visual, and even tactile stimuli. The same examiner, functioning as a control device, can determine what questions are most likely to be informative about a particular examinee. The paper and pencil form of group administered tests has none of these advantages but it is much cheaper. As a result, fixed format, paper and pencil testing is often the method of choice for personnel screening involving large groups of people. Notable examples are tests used for personnel classification in education, the military, and the civil service. Paper and pencil tests have been criticized for their inflexibility and limited scope, but their cost-effectiveness has not been seriously challenged.

Computer-controlled test administration (1) falls between the individual and the group testing situations. Almost any question that can be presented in paper and pencil format can be presented using a computer controlled, television-like display. Furthermore, it is relatively easy to program rules to make the next display contingent upon the response to past displays; i.e. to have the computer assume the control functions of the examiner in the individual testing situation. Finally, test scoring and administration are greatly simplified if responses are recorded directly by a computing system as not others. The evidence is mixed in both cases. Furthermore, it is not at all clear that the change produced by computerization would necessarily be a change for the worse.

Responding to computerized tests evidently does require a specialized ability to deal with the test format. This was shown by Lansman et al. (1982) in a study that examined the correlations between computerized and paper and pencil tests of the "crystalized" and "visual" factors of intelligence that are specified in the Cattell-Horn theory of intelligence (Cattell, 1971; Horn and Donaldson, 1980). Lansman et al. found that the error components of the various computerized tests were correlated with each other, but not with the error components of the paper and pencil tests. This means that there are separate factors for item presentation mode, but that the presentation factors do not interact with other factors. Therefore tests presented in either format would be equally valid tests of intelligence factors, providing that the relative loadings of tests on the "psychological" and "mode" factors do not change across testing conditions. While definitive studies have not been done, there is evidence that this is the case. Computerized and paper and pencil versions of the Raven Matrix test appear to be essentially equivalent (Calvert and Waterfall, 1982; Watts et al., 1982). This is important because the Raven Matrix test is widely considered a good measure of general intelligence.

The studies just cited involved homogeneous groups of subjects. It is possible that certain groups of subjects would be selectively favored or disfavored by computerized testing. The issue of not others. The evidence is mixed in both cases. Furthermore, it is not at all clear that the change produced by computerization would necessarily be a change for the worse.

Responding to computerized tests evidently does require a specialized ability to deal with the test format. This was shown by Lansman et al. (1982) in a study that examined the correlations between computerized and paper and pencil tests of the "crystalized" and "visual" factors of intelligence that are specified in the Cattell-Horn theory of intelligence (Cattell, 1971; Horn and Donaldson, 1980). Lansman et al. found that the error components of the various computerized tests were correlated with each other, but not with the error components of the paper and pencil tests. This means that there are separate factors for item presentation mode, but that the presentation factors do not interact with other factors. Therefore tests presented in either format would be equally valid tests of intelligence factors, providing that the relative loadings of tests on the "psychological" and "mode" factors do not change across testing conditions. While definitive studies have not been done, there is evidence that this is the case. Computerized and paper and pencil versions of the Raven Matrix test appear to be essentially equivalent (Calvert and Waterfall, 1982; Watts et al., 1982). This is important because the Raven Matrix test is widely considered a good measure of general intelligence.

The studies just cited involved homogeneous groups of subjects. It is possible that certain groups of subjects would be selectively favored or disfavored by computerized testing. The issue of motivation is central. Some observers have claimed, more or less without data, that computer controlled tests can be presented in a game-like format, and thus be self motivating. Such an effect could be specific to particular groups of individuals. Casual observation suggests that some people are indeed highly motivated by computer game formats. Others find man-computer interaction intimidating, boring, or both. Intuitively computer games are very much associated with the youth culture (Loftus and Loftus, 1983). Older people may view them with distaste. This is especially true if the testing procedure overtly mo.. res speed of performance, for there is considerable evidence that older people dislike situations in which rapid responding is forced upon them (Hunt and Hertzog, 1982). To further complicate the issue, there is evidence that points the other way. European studies have found that computerized tests are more acceptable to elderly and handicapped people than are pen and pencil tests (Carr et al., 1982; Watts, Baddeley, and Williams, 1982). Given the small data base, it would be inappropriate to conclude anything at all, yet, about the acceptability of computer-controlled testing in different populations.

The issue of motivation is closely related to the issue of prior experience. It is simply not true that "everyone" has or will have had extensive experience with computer terminals during the 1980s. A more realistic assessment is that different people will have very different experiences. Playing arcade games that are computer controlled is not like using a text editor to write term papers. Intuitively, experience with keyboards and terminals would be expected to affect performance in a testing situation. Marshall-Mies et al.

(1983) found a correlation of .2 (in a sample of 500 subjects) between self-report of experience with computer games and performance on a variety of computer-administered tests. Is this so small as to be trivial? The examinees were electrical power system operators who had an average of over eight years experience with computer controlled displays. Clearly the issue is not closed!

In any case, it is well known that virtually any task that involves speeded reactions will show very large learning effects. This raises two concerns. In assessing the comparability of scores from different groups, a frequent issue in psychometrics, care must be taken to ensure comparability of experience with the test apparatus. Within a single group, the presence of large learning effects raises the question of stability in factor patterns. If a battery of computerized tests is given repeatedly to the same people the mean values of the test scores will almost certainly change. Will the test covariances change? Will the correlations between test scores and outside criteria change? These issues are relevant for any testing program, but may be particularly important for any computerized testing that involves the measurement of reaction time.

One of the strongest arguments for computerized testing is that it makes possible adaptive adjustment of test difficulty to the examinee's talents. It is conceivable that such a testing procedure could interact with personality traits. Consider an individual who is of slightly below average talent. Adaptive testing would begin by presenting this person with items of average difficulty, which will prove too hard. The examinee will thus be given an initial series of failure experiences as the computer attempts to find an appropriate level of item difficulty. Now suppose that the individual is also test anxious, a well identified personality tra' (Sarason, 1980). Will the failure experiences early in the test cause the person to under perform on the remaining items? We simply do not know.

There are two issues concerning "testing in the abstract" that revolve around the use of test results more than around the interaction between the examinee and the test administrator. Tests of cognitive ability are typically used for one of two purposes; personnel classification or diagnostic evaluation. In personnel classification the purpose of testing is to identify individuals who have (or lack) some degree of specific ability, so that these individuals may be offered or denied an assignment. College entrance examinations and Armed Services screening examinations are examples of classification tests. Classification testing is usually oriented toward the examination of broadly defined ability domains, such as verbal, spatial or reasoning skills. A relatively brief amount of time is devoted to gaining in-depth information about a person's ability within one of these domains. There is extensive predictive validity data showing that there is a correlation between test scores obtained in classification testing and similar broad criterion variables, such as academic achievement or success in service technical training schools (Burns,1975; Jensen,1980). In assessing the validity of a classification method the appropriate criterion is a population one. Does the testing procedure justify itself by providing data that reduce the costs of misclassification below the cost of the test <u>on a population basis?</u> Computerized tests have to show that they can meet this criterion, in competition with other methods of testing. Put another way, suppose that the correlation between a computer administered test and a paper and pencil test was 1.0. Then one should immediately decide to use the method of testing with the lowest administrative costs. This will seldom be the computer administered test.

The purpose of diagnostic testing is to gain a detailed knowledge of the pattern of performance of a specific individual. Absolute levels may be more important than levels relative to population norms. Neuropsychological examination is a prototype case; the purpose of the examination is not to select a person for entrance into a fixed environment but to design an environment for the individual. Other, less extreme examples can be envisaged. The scores to be obtained and the criterion variables to be utilized will vary from case to case. The introduction of computers into diagnostic testing thus may pose a much more difficult problem in validation than is the case for classification testing. On the other hand, the flexibility of computer administrated testing should provide a greater potential in the diagnostic than in the classification area.

Matarazzo (1983) raised a somewhat different issue. He was concerned that the widespread availability of computer-administered tests of intelligence and personality might lead to misapplication of the results. Matarazzo focused upon misinterpretation of and over-reliance on tests by naive users, simply because the tests were computerized. These issues, while real, are essentially issues that concern psychology as a professional guide rather than issues that are directly relevant to psychology as a science. Concerns about oversell because something is on a computer are not qualitatively different from concerns about over-acceptance due to some other irrelevancy, such as endorsement by a famous athlete or entertainer.

In discussing the scientific potentials for using computers to test intelligence, we must first define what it is that we are trying to test. Initially we shall take the Boring (1923) view that intelligence is defined by the intelligence test. Accordingly, we should first consider how computers can be used to expand upon current testing of those psychological functions that are tested by present day methods. Such expansions are steps forward. We then take a leap into new areas, bu suggesting some psychological functions thatt are logically part of mental competence, but are not tested by present day methods. In discussing both steps and leaps we have attempted to distinguish between questions of test development, given the present state of knowledge, and questions that require further psychological research.

Modification and Extension of Present Assessment Procedures

Although there are many taxonomies of human mental performance there is also considerable agreement about the major intellectual dimensions that must be assessed in order to evaluate mental competence. Three types of "intelligence" will be considered; verbal ability, spatial-visual ability, and logical reasoning. While varying studies have found these abilities to be moderately correlated in most populations, they clearly represent distinct aspects of human thought (Carroll. 1982). We will also consider the evaluation of memory, as memory assessment is part of individually administered test batteries.

Verbal Ability

Modifying current assessment. "Verbal ability" refers to competence in comprehending linguistic messages. In order to exercise verbal ability on needs skills that go well beyond knowledge of word meanings. Nevertheless, on a statistical basis, people's ability to comprehend stories, utilize grammatical rules, and do other linguistic tasks are all well predicted by performance on a vocabulary test. (Carroll.1979). Put another way, if we want to know what verbal ability is, one has to consider mmany things besides word knowledge. If all that is desired is an estimate of a persons's verbal facility, a vocabulary test will often suffice (Carroll, 1971, Hunt, in press).

One of the open questions in intelligence testing is whether people differ in the size or in the nature of their vocabularies. The opinions of experts differ dramatically. Nickerson (2) has stated that people differ in their vocabularies by a factor of as much as two to one. On the other hand, Fillmore (1979) has stated that people do not differ very much in their total vocabulary size, but to differ in the sort of vocabularies they have. Neither Nickerson or Fillmore provide citations for their claims. (Carroll, 1971) cited data indicating that there are large individual differences on common vocabulary tests but thisdoes not settle the issue. In the traditional intelligence test a fairly small set of words is designated as "representative" of word knowledge in general, and the examinee is tested on this set of words. Such a procedure implicitly accepts the idea that most of a person's vocabulary is drawn from a universe of standard words. If testing were done by a computer it would be possible to probe knowledge of specialized vocabularies, by drawing test items from a population of words chosen to reflect the examinee's presumed characteristics. To take a simple example, if an examinee claimed to be a native speaker of Spanish it would be possible to probe for word knowledge about cars, if the examinee claimed expertise as an automobile repairman, or for word knowledge about music, if the examinee claimed to a modern music fan. This procedure will be called tailored vocabulary testing.

Tailored vocabulary testing can separate the issue of the size and nature of a person's vocabulary. This could be important. Acquisition of a vocabulary depends upon two things: culturally determined exposure to a particular lexicon and personal sensitivity to verbal stimuli in general. Logically these are separate traits. Present tests, by measuring standard (American) English vocabulary, are undeniably biased toward measuring cultural exposure. For this reason it can be claimed that the tests are "unfair". If so, why do vocabulary tests serve as useful predictors? It could be that, in spite of their cultural bias, the tests still depend enough upon individual verbal facility to be useful measures of the underlying psychological trait. On the other hand, it could be that in some circumstances the tests predict future behavior precisely because they measure cultural background (3). Arguments could be made for either conclusion. A testing approach that evaluated a person's use of both standard and specialized vocabularies would make it possible to investigate the issue further.

<u>New Directions.</u> The co-ordination of verbal and non-verbal input can be looked upon as a prototypical verbal action. Tasks that require the action will be called "co-ordination tasks". An example is the comparison of verbal descriptions to directly perceivable visual, auditory, or tactile displays. Following instructions often requires co-ordination; e.g. "Turn the knob t the left if the red light goes on." There are substantial individual differences both in the strategies that people use to co-ordinate verbal and non-verbal input, and in the effectiveness with which they execute those strategies (Hunt, 1980; in press). It appears that the sort of "verbal ability" tested by co-ordination tasks is partially independent of "verbal ability" as tested by tasks requiring only that people respond to isolated words (Hunt, Davidson, and Lansman, 1981).

Most of the experimental paradigms used to study the co-ordination of verbal and non-verbal information rely on computer presentation. Physically, it would be easy to adapt them for psychometric testing. Two examples are the <u>sentence verification and instruction following</u> paradigms. In sentence verification the examinee must decide whether or not a verbal statement correctly describes a picture. The complexity of the verbal statement correctly describes a picture. The complexity of the verbal statement can be systematically manipulated (Carpenter and Just, 1975). Testing the ability to follow instructions, the examinee is told to execute actions contingent upon presentation of certain displays. Again the variable of interest is the examinee's response to changes in the linguistic complexity of the instructions, rather than to changes in stimulus or response complexity itself (Dixon, 19[illegible]). The instruction following paradigm is of interest because it tests the comprehension of verbally stated procedures. Most present day verbal tests are of the comprehension of expository material.

There are a number of practical psychometric issues to be resolved before either of these paradigms could be introduced into testing situations. They revolve around test-retest reliability, across time periods and across likely variations in testing situations. Such questions fall somewhere between "basic" research and "developmental studies", but they must be answered before we know whether or not it is practical to test people's ability to do co-ordination tasks. SImilar points will arise frequently in our subsequent discussion.

Several computer-controlled procedures have been developed for use in the study of verbal comprehension process in general. Some of these involve highly intrusive measurement devices, such as instruments for recording eye movements. These are probably not practical testing devices. A number of other laboratory paradigms involve what is called the <u>secondary task</u> methodology. The mental effort required to execute a linguistic task (e.g. parse a passive sentence) is measured by the ease of execution of an ancillary simple task, such as phoneme monitoring. Individual differences can be evaluated using the secondary task paradigm, but doing so requires an elaborate experimental design that is probably not practical in most testing situations (Hunt and Lansman, 1982).

Rapid single visual presentation (RSVP) is a technique for reading in which readers comprehend a message that is presented visually, one word at a time (Potter, Krull and Harris, 1980). The rate of presentation can be controlled either by the reader or the examiner. To our knowledge the RSVP technique has not been used to evaluate individual differences in the dynamics of comprehension such as changes in the rate of reading as a function of te complexity of texture material. It might be that the method could be adapted for computerized testing, but a great deal of work remains to be done to establish appropriate testing procedures. The RSVP technique seems particularly appropriate in situations in which the diagnosis of reading style is an issue.

Conclusion: Computer controlled testing is not likely to extend the range of evaluation of verbal ability in classification situations; if only because the simple vocabulary test is such a good predictor. Therefore any justification for computerizing verbal classification procedures will have to rest on cost effectiveness. Computer controlled testing may prove much more useful in diagnosing individual comprehension skills. Such testing may be quite valuable if there is reason to believe that the examinees have been exposed to some condition that would affect the non-lexical aspects of verbal processing. For example, a variety of reports have indicated that age deleteriously affects the attention demanding, speeded aspects of verbal processing, although vocaulary size probably increases with age (Hunt and Hertzog, 1981). Specific reading disabilities in schoolchildren involve selective loss of the ability to deal with particular aspects of written material (Vellutino, 1979). Even more dramatic selective losses of function are sometimes seen in cases of brain injury. Prototype computer-controlled testing procedures for various aspects of reading comprehension have been developed (Frederiksen, 1982) and could be extended to more general situations where large scale, rapid diagnosis of language deficiencies were required.

Visual-Spatial Ability

Modifying Current Assessment. The term "spatial ability" refers to the ability to construct, examine, and manipulate internal representations of visual displays. Examples of standard spatial ability tests are the Minnesota Paper Form Board test, the Primary Mental Abilities 'spatial' test, and the KohsBlock Design test (McGee 1979). The well known "mental rotation" paradigms developed by Shepard and his collegaues (Shepard and Cooper, 1983) have also inspired some more specific tests of spatial ability (Lansman et al., 1982; Vandenberg and Kuse, 1978). The domain of abilities covered by spatial tests is reasonably well fit by a two factor model (Egan, 1979). One factor, "spatial encoding", is an ability to recognize a specified shape when it occurs in the visual field. Examples of encoding tasks are recognizing letters in a word or recognizing the triangles embedded in a six pointed Star of David. The manipulation of a metal image is a rather different ability. It is best illustrated by the rotation tasks developed by Shepard and his associates. A more prosaic example is the mental rotation of letters when reading a text that is presented upside down. We will refer to the aility to manipulate images as "visualization", although this is perhaps not the best term.

Computerized testing methods offer a considerable potential for expanding the evaluation of spatial ability than in the verbal domain. Most preent spatial-visual tests mix the encoding and visualization factors. Pellegrino and his associates (Mumaw and Pellegrino, in press; Pellegrino and Kail, 1982) have shown that it is possible to obtain pure measures of both the visualization and encoding factors within the framework of a single test. This is done by presenting successive complications upon a single base problem (e.g. a form board completion item) in ways that increase the difficulty of the base problem along either the encoding or visualization dimension. The item construction procedures developed by Pellegrino et al. could easily be combined with the response analysis procedures based on multicomponent latent trait theory (Whitely,1981) and upon adaptive testing. This is an example of how the computer technology offers a chance to make a solid step forward in our evaluation methods.

New Directions

Our present ideas about spatial ability are based upon data that can be gaterhed using the fixed format, paper and pencil testing technique. The paper and pencil format is much more restrictive in the spatial than the verbal domain, because the mapping between the dynamic, three dimensional, colored world and the two dimensional, static, (largely) black and white printed diagram is not as tifht as the mapping between listening and reading. Hence there is a real potential for expanding our concept of spatial ability by adopting a new testing format. In particular, computer controlled displays can be dynamic. This is an important distinction, because spatial abilities are thought to be involved in tasks in which motion is an inherent part of the environment, e.g. operating a vehicle or an aircraft. It would be relatively easy to test the ability to perceive and extrapolate motion directly by using computer controlled animated displays. Is the ability to deal with motion really predictable from our present static tests of spatial ability? If not, would measring it add to our ability to predict performance in tasks that we beleive require spatial ability?

Traditional spatial ability tests deal with the ability to look at a display. What about the ability to orient oneself to the surroundings? We will refer to this as "geographic orientation." Although there is evidence that people vary widely in their ability to maintain geographic orientattion (Thorndyke and Goldin, 1981) there is little systematic knowledge of the dimensions of the ability. The paucity of data may be due to the sheer economics of testing. Studies of geographic orientation require that people move about in the environment. Is it possible to conduct geographic orientation studies less expensively, by using computer controlled technology? People could be provided with the visual input from imaginary walks or drives, by combining computer controlled testing with te se of a random access video disk device. They could then be tested on their knowledge of the geography that they had experienced vicariously. Would the same people who did well in reconstructing the simulated geography do well at maintaining their orientation if they were moved through a real environment? The experiments need to answer this question are again straightforward. The answer itself s important because of its implication for future research. If orientation in the simulated environment correlates with orientation in the actual environment, then performance in the simulated environment ould be used either as a predictor of actual geographic orienttion ability, or as a means of studying the relation betwenn orientation and conventional measures of spatial ability.

The above examples have concentrated on the "spatial" part of spatial-visual ability. Gopher (3) has pointed out that many (though not all) of the classical visual illusions can be tested using computer generated displays. Several of these illusions are thought to indicate basic properties of higher order perceptual systems. To illustrate, Coren and Girgus have shown that there are individual differences in susceptibility to various perceptual illusions, such as the Muller-Lyer illusion. The dimensionality of the space of susceptibility is considerably less than the number of illusions. Some of the factors of susceptibility seem to be related to peripheral visual processes while others seem to be associted with higher order, cognitive factors. Very little is know on how susceptibility to illusions relates to other dimensions of intellectual or personality variation. (A striking exception to this statement is the very large literature on field dependency and rod-and-frame illusion (Witkin et al., 1954). If computer generated illusions can be produced reliably this could be a fruitful field of research. (4).

pm+0

Conclusion : The advent of computer controlled testing should considerably expand our ability to examine spatial-visual abilities.

Enough research has already been done to warrant beginning the costruction and validation of computer-oriented tests. Studies intended to expand our conceptualization of spatial ability to include dynamic visualizations and geographic orientation could have an extremely high pay off. The extension of testing into the realm of perceptual illusions is more venturesome, but is worth considering.

Reasoning

Modifying Current Assessment. Reasoning is traditionally divided into two areas; deductive and inductive reasoning. Deductive reasoning requires that a general principle be applied to a particular case, whereas inductive reasoning involves extraction of a general rule or principle from examination of cases. Tests of reasoning have been widely used in the study of intelligence. Since the time of the ancient Greeks, deductive reasoning has been seen as the hallmark of rational thought (Johnson-Laird, 1983). Inductive reasoning (including analogical reasoning) has been shown to be among the best markers for the general ("g") factor in modern intelligence tests (Sternberg, 1977). This is intuitively reasonable, for a good argument can be made that the essence of intelligence is being able to transfer problem solving techniques learned in one situation to other, analagous situations.

During the 1970s and early 1980s considerable progress was made in understanding the psychological basis of both inductive and deductive reasoning. There are now several, rather similar models of how people represent various types of reasoning problems to themselves, and of how those representations are manipulated during problem solving. (Goldman and Pellegrino, 1984; Johnson-Laird and Steedman, 1978; Rips, 1983, Sternberg, 1977,1980,1982). These models provide good explanations of why particular reasoning problems are easy or hard, and they identify microprocesses that must be executed in order to solve an inductive or deductive problem. The models could be used to design reasoning problems that vary systematically along known psychological dimensions, in much the same manner as was suggested for spatial tests. Obviously, computer testing techniques could be used in such an effort, but the effort itself does not require computers (see e.g., Holzman, Pellegrino and Glaser, 1983). The cause for this conclusion is worth examining. Reasoning is, by definition, the task of operating upon an internal, conceptual representation of a problem. The computer provides a facility for displaying different problem formats, but this should only affect the process of constructing the conceptual representation, not the process of manipulating it. Reasoning is also a process that often takes place over minutes rather than milliseconds, and our interest typically centers on errors as much as speed of reaction. While reaction time measures can be useful in testing models that identify components of solution during reasoning tasks, (Sternberg's, 1977) it is unclear whether or not reaction time measures of component processes of reasoning are sufficiently reliable to be useful in brief tests of a person's reasoning power (Goldman and Pellegrino, 1984) Thus the contribution of computerization to classification tests of reasoning is problematical.

Computerized testing may have more of a place when testing is done for diagnostic purposes. Reasoning problems are typically attacked in stages; encoding the elements of a problem, applying trial solutions, and evaluating solutions. Different strategies stress one or the other of these stages. Furthermore, success in problem solving appears to be related to the choice of a strategy. Sternberg and his associates have consistently found that good problem solvers spend relatively more time reading a problem and understanding the elements involved, whereas poorer problem solvers seem to begin trying out solutions (Sternberg, 1982). Dillon and Stevenson-Hicks (1981) have reported differences in problem evaluation. When good problem solvers take multiple choice reasoning tests, they examine the problem, develop a solution, and then look for the solution in the set of answers. Poorer problem solvers spend more time examining and rejecting the alternatives provided in the answer set. Note that the former strategy generalizes beyond the multiple choice test format, while the latter strategy does not. Finally, Egan and Grimes Farrow (1982) have shown that peoples who adopt an abstract public representation are more successful in solving transitivity problems than are people who adopt a concrete representation. A variety of interactive computer controlled testing procedures could be used to identify a person's style of problem solving, and perhaps to arrange ways of improving on it. But is this a good idea? The research cited shows that different problem solving styles exist. It does not show how stable a person's problem solving style is over different occasions, or different situations. The problem of situational generality is of particular interest.

New Directions. From the viewpoint of a logician, the rules for inductive and deductive problem solving apply regardless of the content area. Psychologically content is far from neutral. Sternberg (1980) has pointed out that there are two stages of problem solving; knowing abstract problem solving methods and knowing when to apply them. The ability to realize that a probl... olving method is appropriate may be much rarer than the ability to apply a method in a familiar area. For example, mental retardates can be trained to apply problem solving procedures, but the application is tightly tied to the context in which it is learned (Campione, Brown, and Ferrara, 1982). Cross cultural examples are also plentiful. Indeed, it seems that the idea that there are abstract reasoning procedures, apart from a particular context, is a peculiarly Western idea, that is acquired through contact with formal Western schooling (Scribner, 1977).

Sternberg (1981) has recently advocated testing people's ability to reason, in the abstract, by testing their ability to deal with problems presented in what are almost bizarre settings. An example is a problem about people who grow young instead of old. He has referred to the ability to deal with such situations as the ability to reason in "non-entrenched" situations. One can envisage a computer-controlled testing situation in which a person's reasoning ability was examined both in settings with which the examinee was familiar and in more foreign settings. A related procedure would be to examine people's ability to use hints to suggest fruitful analogies in problem solving. This technique could be used to provide a finer analysis between individuals than the discrimination between 'solvers' and 'non solvers.'

Virtually every discussion of reasoning closes with a rather ill defined section on 'problem solving.' Computer programs can be written to present quite challenging problem solving tasks-e.g. the ubiquitous 'dungeons and dragons' games advertised extensively in computer journals. Could these games be used to test problem solving skills? Although the idea is appealing, the idea does not always work. Time to solve a realistic computer-presented problem in electric power distribution correlated only -.16 (N=3441) with supervisors' ratings of power system operators (Marshall-Mies, 1983).

This discouraging result can be used as an illustration of a prosaic but important point. The concerns about test length and item reliability that are routinely dealt with in conventional testing are also concerns in computer-administered testing. If a test consists of only a few items, as in Marshall-Mies et al.'s use of a single problem, then it is unreasonable to expect that the test will be reliable. On the other hand, if there are many items, it is possible to produce a very reliable test even though each item has a substantial item-specific component (Green, 1981). Since reasoning and problem solving tasks are frequently found to have a considerable amount of item-specific variance, extended testing is essential if one wants to evaluate a person's general problem solving or reasoning ability. To drive this point home, consider a study by Sternberg and Gardner (1983), that reliably evaluated people's ability to execute particular components of a reasoning strategy. Such an evaluation could be regarded as the ultimate goal of diagnostic testing. Sternberg and Gardner's participants solved over one thousand problems each! This illustrates how serious the problem of item reliability is.

Conclusion. Can computer-controlled testing help by speeding up item presentation and response evaluation? Speeding up these functions of test administration would be only marginally useful. The limiting factor on speed of reasoning is almost always the speed with which people think, not the speed with which the problem is presented to them. Since the bottleneck is mental processing speed, speeding up test administration is almost irrelevant If there is to be a synergism between computerized testing and the evaluation of reasoning, it will be based on a sophisticated method for diagnosing problem solving strategies.

Memory

Memory evaluation has always been part of individually administered psychological tests. The Wechsler Adult Intelligence Scale (WAIS) contains a memory subscale and a memory quotient score in addition to the more familiar intelligence quotient.(Matarazzo, 1972) In spite of their ubiquitousness, the memory scales used in psychometric tests are simplistic compared to the measures of memory that are routinely utilized by experimental psychologists and neuropsychologists. The WAIS scale itself is inadequate to differentiate among types of amnesia, even though the instrument is frequently used to diagnose the amnesic syndrome (Hirst, 1982).

The problem seems to be that the psychometric measures are, for the most part, based on a naive view of human memory. Conventional memory subscales focus on the global functioning of two memory systems, very long term ("permanent") memory for facts; and memory span, the capacity to recall, exactly, material presented within the last minute or so. A more modern view of memory requires testing of at least three systems; the two described above and a third, working memory, that contains a general representation of the current situation without necessarily containing an exact record of all details of the information that has been presented (Baddeley, 1976; Hunt, 1977; Klapp, Marshburn, and Lester. 1983). Furthermore, most experimental psychologists would be unsatisfied with simple measures of capacity. In addition they would like to have a measure of the speed of retrieval from various memory systems and some measure of the rate of transfer of information between systems.

Procedures for estimating such parameters are routinely used in experimental psychology today. Many of the procedures demand the sort of control that can be achieved only by computerized testing. Furthermore, at least some of the measures are related to memory functioning in extra-laboratory situations. (Sunderland, Harris, and Baddeley, 1983). Thus there is clearly a potential for expanding present methods of assessing memory in a way that would not be feasible using conventional testing methods. However, more is required than simply adding experimental paradigms into a testing battery. The paradigms generally have been designed to estimate group parameters. Whether or not they are sufficiently reliable to be used on an individual basis remains to be shown. Furthermore, most of the experimental paradigms demand much more time from an individual subject than would be available in a classification testing situations. Finally, a person's apparent memory capability may depend heavily upon a person's familiarity with the content of the information being memorized. It appears tht people will develop content-specialized frameworks for dealing with information that they encounter repetitively. To take an extreme example, Chase and Eriksson (1981) report the case of an individual who had a memory span for numbers in excess of eighty (!), because he had developed a special technique for memorizing numbers. His memory span for other material was in the normal range.

Results such as Chase and Eriksson's pose a serious question about the evaluation of memory. Is memorization ability a stable characteristic of the individual, or is it a product of the interaction between the individual's capabilities and his or her specific experiences? Theorists are clearly divided on this point. Consider such a basic point as the development of short term memory. Some have argued that the development of an ever increasing short term memory capability is the cornerstone of cognitive development. Others have argued that the apparent growth in memory is due to growth in the child's knowledge of memory strategies and due to a widened familiarity with useful, content-specific codes. (Siegler and Richards, 1982). The controversy is important in its own right. It is important to the development of computerized testing, because the testing procedure should be based on a clearer idea of what is being evaluated.

Posing the issue as a contest between theories of 'generalized memory ability' and 'situation specific memory' is too simplistic. It is clear that if clinically normal individuals are given special training, so that they develop 'expertise' within a content area, then their ability to remember information in that content area will be selectively improved. However, it is possible to develop tests of unfamiliar expository or arbitrary material. Most of the paradigms developed in the experimental laboratory are of this type. A substantial study of a performance on such tasks in a college student population (Underwood, Boruch, and Malmi, 1978) found that a very large percentage of the variance on all tasks could be accounted for by a general associative memory factor. This finding suggests that one or two memory tests would suffice to estimate general memory performance in a variety of situations.

The problem is thornier when 'exceptional' memory is to be assessed. The issues involved in evaluating exceptionally good or exceptionally bad memory are quite different.

Extremely poor memory is usually associated with some pathological condition, such as Alzheimer's disease, senility, or brain injury. Testing the affected individuals is necessarily a form of clinical interview. Computer-controlled presentation provides a chance to standardize the interview, which has some advantages, but the testing procedure will probably always be an adjunct to individual patient assessment by a clinician. The role of computer controlled testing in the assessment of superior memory is rather different. As has already been pointed out, computer controlled testing is flexible, so that within a fixed time it would be possible to test a given individual's memory for more different content areas, than could be evaluated using conventional testing. The remarks made earlier about tailored vocabulary testing are applicable. Thus computerized testing would be useful in identifying the fact that a person had superior memory for certain types of materials.

What would be hard to do is to establish why an unusually good memory capacity existed. Superior memory capacity is generally associated with the use of elaborative strategies that are appropriate to the material being memorized. (See Bransford et al., 1982 for a good discussion of this point and for a list of further citations.) There are a few situations in which strategies can be inferred from computer-analyzable responses, such as reaction times (MacLeod, Hunt, and Mathews, 1978) or eye movements (Dillon and Stevenson-Hicks, 1981). In general, though, the best way to evaluate a person's memory strategy is to ask the examinee to describe it, and then have an experimenter analyze a complex verbal response. While advances in computer comprehension of freely produced speech have been made, it seems unlikely that computers will be prepared to replace the experimenter in the near future.

New Dimensions of Testing

Learning, attention, and psychomotor skills are certainly aspects of mental competence. However, they are not usually evaluated in intelligence testing. Does computerized testing offer a chance to expand evaluation into these fields?

Learning

Definitions of what "the intelligent person" can do almost always include learning (Sternberg et al., 1981). Intelligence tests make relatively little effort to assess learning directly, because of the difficulty of doing so using a conventional test format. Since many of the paradigms used in laboratory studies of learning now utilize computer-controlled experimental procedures, the advent of cheaper computing offers the technological possiblity of assessing an individual's rate of learning. For example, it would be possible to observe a person's learning rate in situations as different as learning to recite lists presented via computer or learning to play a video arcade game.

Experimental studies of the feasibility of such testing have already been reported as part of an attempt to develop new military classification procedures. (5) Such proposals presuppose that there is such a learning ability apart from the ability to learn within specific content areas. The evidence for this proposition is far from clear. The assesment of learning ability raises a number of issues that transcend the narrower issue of computerized testing. It is instructive to see how the larger issues and the specific ones interact.

Types of Learning. Estes (1982), in a perceptive review of the issues, has pointed out that learning situations can be ordered by complexity. At the 'simple end' are studies of habituation and conditioning. Instrumental learning (operant conditioning) and paired-associates paradigms, are more complicated, while concept identification and verbal learning paradigms can be very complex. At the extreme, one could regard any education as a form of learning. Here the distinction between learning and induction becomes arbitrary.

The simpler learning paradigms, such as habituation and conditioning paradigms, attempt to isolate the formation of elementary stimulus-response bonds. As Estes (and others) have pointed out, studies of extreme contrasts as the contrast between mildly mentally retarded and normal individuals have not shown particularly striking relations between performance in simple learning situations and measures of general intellectual competence. Thus such paradigms will not be considered further.

The problem of evaluating a person's ability to learn complex material is complicated by three factors. When new facts are learned they are fitted into a person's existing information. Therefore any evaluation of a person's ability to learn material with meaningful content will interact with the examinee's prior knowledge. A determination that a person is a rapid learner of material in history may have little predictive value for estimating the person's ability to learn auto mechanics. Since learning is highly dependent on memory, the remarks about individual differences in memorization strategies apply. Because individual differences in knowledge and strategies of information acquisition interact, a person's learning ability may best be evaluated by constructing a model of how the learner conceptualizes the learning problem. There have been attempts to develop computer programs that do this, by observing students' responses to computer controlled presentations of material to be learned (Brown, and Burton, 1978; Collins, 1975; Sleeman, 1983).

Considerable research needs to be done before these programs are sufficiently validated so that their results could be used in personnel selection or diagnosis. In fact, the time required to induce a model of the learner may rule out the use of such student evaluation devices in classification situations. The practicality of using learning modeling programs for a diagnosis of the learner's knowledge has not yet been established, but is a potentially interesting research topic. The above remarks apply to studies of "learning ability" as evaluated within a single testing session. An alternative method of testing is to embed systematic assessment within computer assisted instruction programs, so that evaluation and learning proceed hand in hand. If this can be done the opportunities for testing increase enormously. This approach also has the advantage of tying assessment directly to the learning of material that presumably has value in its own right. Those instructional programs that attempt to develop models of the learner (see above) could be regarded as sophisticated teaching and assessment situations.

<u>Procedural Learning.</u> These somewhat discouraging conclusions are based largely on studies of the learning of verbal or quasi-verbal material. Learning also takes place in situations in which the task is to classify and respond to perceptual displays. Here the extra-laboratory analog is to machinery operation, survelliance of radar screens, or even to athletic performance. Studies of experts in

such fields indicate that a great deal of their performance is carried out in what has come to be known as the "automated mode" of performing. Reactions to displays are overlearned and immediate. This is contrasted to the slower and attention demanding "controlled" mode of responding. The somewhat intuitive distinction between automated and controlled responding has been operationally defined by Schneider and Shiffrin (1977), in a series of experiments on the development of automated responding to visual displays. They found that automated responding could take literally thousands of trials to develop, but that when it did develop reponses might be as much as an order of magnitude faster than reponses made in the controlled mode. Subsequent studies (Fisk and Schneider, 1983; Poltrock, Lansman and Hunt, 1982) have shown that the automated-controlled distinction applies to a variety of item recognition situations, quite outside the narrow area of visual display recognition. Since the techniques for evaluating automated and controlled reponding require very rapid presentations of stimuli and recording of responses, some form of computer control is necessary. Could the experimental procedures be adopted to testing situations?

There are two basic questions that one would ask in studying individual differences in automated and controlled responding. Given a number of candidates for training on a task, which candidates are likely to develop automated responding early in training? Given a number of people who have already been trained to execute the task, which persons have developed to the stage of automatic responding? The first question is more interesting from the viewpoint of personnel selection, but unfortunately it may be the more difficult to answer. The straightforward way to ask the question would be to present candidates with a (computer controlled) simulation of the task to be learned..or perhaps with the task itself..and observe changes in performance over a relatively small number of learning trials. If learning rates early in training predict learning rates throughout training the approach should work. But do they? Here the evidence is mixed.

Schneider (Note 6) has presented data showing that learning performance early in training on a task analogous to interpreting a radar display is not predictive either of asymptotic performance or of the time required to learn an automated skill. Schneider studied a simple visual scanning task, analagous to the interpretation of a display on a radar screen. A somewhat similar design was used by Dunlap, Biladeau, and Jones (in press), but in their case the tasks were commercial video games that appear to involve visual scanning, target identification, and motor reactions. The factor structure of the tasks changed very little over extended practice.

It is relatively easy to determine whether or not an individual has been able to automate performance on a specific task. Automation is defined by insensitivity of performance to increases in information processing loads. Computerized evaluation procedures could be developed to monitor the stage of a person's training in a variety of situations in which automation was important. The evaluation procedures would be specific to the task at hand, rather than being general indices of individual trainability. In general, the automated-control distinction seems to be more relevant to evaluating the success of a training program, than for predicting the success of a given individual entering that program.

<u>The Proximal Zone of Development</u> The above remarks suggest strongly that learning is situation specific. Therefore it makes sense to build assesment into specific computer controlled teaching programs, but it makes less sense to try to develop comprehension procedures for evaluating learning ability in general. There are those who would dissent with this viewpoint. The dissenters believe that there is a general learning factor that applies to more cognitive situations, such as learning how to solve problems in mathematics or reading. A central idea in their view is the concept of a person's "proximal zone of development." Loosely, people who have wide proximal zones are people who are able to utilize hints to learn how to solve unfamiliar problems. The concept was originally introduced by Vygotsky (English translation 1978), in the 1930s, and appears to

have influenced Soviet personnel evaluation. The idea, although not the language, is similar to Sternberg's (1981) notion that one should distinguish between a person's ability to do familiar (entrenched) tasks and his or her ability to deal with novel, non-entrenched problem solving. Outside of the Soviet Union, Vygotsky's idea has been translated into action in a number of programs in which children's potentials are assessed by giving them progressively more hints until they can solve test problems. The idea is that the child is simultaneously being trained to be a problem solver and being evaluated for the ability to deal with novelty. The word "child" is used advisedly here, as all the programs that we are aware of are directed toward children, and often abnormal children at that (Campione et al., 1982; Feurstein, 1980). At present the task of evaluating a person's potential for learning is conceived of as an almost clinical interview, involving intense social interaction between the examiner and the person being evaluated (this is especially the case in Feurstein's work.) Recently, though, some of the testing procedures have been adopted for computer administration (Campione, Note 7).

The apparent contradiction between the work of Feurstein and Campione et al., on the one hand, and the other studies cited may be due to the population tested. It could well be that there are significant individual differences in generalized learning ability amongst slow learners, while learning might depend more on specific prior knowledge in normal and rapid learners. This important issue is not directly related to the issue of the use of computers to assess learning ability. The answer to the question has strong implications for the design of computer programs intended to assess that ability.

Psychomotor Skills

Intuitively, it makes sense to speak of people who are clever with their hands. On the other side of the coin, the evaluation of psychomotor skills is almost completely absent from current psychological testing. In part, this may be because it is hard to include a test of tool manipulation in a standard paper and pencil testing situation. One could imagine fairly simple psychomotor tests that could be administered by a computer. The simplest example would be some form of tracking using an electronic "mouse", a light pen, or a similar pointing device. Other tests of tool use could no doubt be devised. Would there be any point in doing so?

Psychomotor skills tend to be highly specialized and to be very sensitive to practice (Newell and Rosenbloom, 1981). Computer games that require psychomotor skills exhibit low session to session reliability until after ten or fifteen days of training (Kennedy, Carter, and Bittner, 1980). Finally, it has long been known that the relative importance of different components of a psychomotor skill may change with the state of practice (Fleishman, 1967). These findings all suggest that there is no general factor for psychomotor co-ordination. If this is true, then the fact that psychomotor tasks can be administered by computer is somewhat irrelevant to testing except to evaluate the results of specific training sequences.

There is a possible exception to this somewhat discouraging conclusion. It has been claimed that the speed of a person's cognitive decision processes can be revealed by the speed of simple discriminations and/or response choices. Examples are the much used choice reaction time paradigm, in which a person must move one of N switches when a light is shown over it and a variety of measures of speed of retrieval of information for long term memory. These tasks require computer presentation because of the need to control stimulus presentation and record response times at millisecond accuracy.

Should the speed of simple but rapid tests be included in a screening battery? The answer to this question depends upon two familiar psychometric issues; stability and external validity. The stability of choice reaction time duties, across sessions, is rather low (approximately .6). This contrasts with their extremely high (.95) intra-session reliability. Such a discrepancy indicates that choice reaction time measures may be subject to intra-subject influences that vary from day to day. Whether day to day variations in a person's reaction times mirror variation in other cognitive behavior is simply not known. (The answer to this question would be interesting). Given the fact of inter-session variation, it is somewhat surprising to find that the correlation between choice reaction times and more complex tasks is as high as it is.

Several studies have shown that choice reaction times are reliably correlated with tests of the "g" factor of conventional intelligence testing (Hunt, 1978; Jensen, 1982; Vernon, 1983). We caution, though, that the correlations (.2 to .4) are not sufficiently high to argue that this test should be used as a marker of general intelligence. Much less is known of the relation between complex behavior and perceptive discrimination speed. There is a sketchy report of a very high correlation between measured intelligence and speed of visual discrimination (Brand,1981), but subsequent work has failed to replicate the original study (Nettlebeck and Kirby,1983, Vernon, 1983). The problems of controlling for simple sensory deficits, such as astigmatism or myopia, have apparently not been studied.

O ly a weak case can be made for using choice reaction time and perceptual discrimination speed as measures of psychological func ons that are also measured by present day psychomet ic tests. In part this is a problem of definitions. If one is willing to accept, say, the Raven Matrix test as a definition of general intelligence, why replace the matrix test with another test that either takes as long, or has only a moderate correlation with the accepted test, or both? A more interesting approach is to regard information processing speed as an extension of the abilities now tested. If it could be shown that adding the measures to a test battery improved the predictability of performance in some class of complex task then an expansion of testing would be appropriate.

Attention

Spearman (1927) suggested that the general factor in intelligence might in part be due to pervasive individual differences in one's ability to concentrate attention on the problem at hand. Several other writers have echoed Spearman's comment. For instance, Matarazzo (1972) asserted that in normal adults the chief use of the memory span test on the Wechsler Intelligence Scale was to test the examinee's ability to concentrate. In spite of such claims, early writers on intelligence offered little empirical evidence relating "intelligence" to "attention." The failure to develop the idea was partly due to the difficulty of monitoring a person's attentional effort, given the constraints of conventional testing situations. It was also partly due to a failure to be precise about what was meant by "attention."

Since 1970 the concept of attention has been articulated more clearly. The most common view is to look upon attention as a resource, analagous to electrical power. Some authors argue for a single, pervasive "power source" for thought (Kahneman, 1973), while others have presented models that postulate two or more sources of non-interchangeable resources (Navon and Gopher, 1979; Wickens, 1979), or competition for limited capacity information processing mechanisms (Hunt, 1981; Posner, 1978). In all cases, though, resource expenditure is measured by observing patterns of interference between two tasks. Consider two tasks, x and y, both of which demand attention. Examples are pursuit tracking and monitoring a stream of signals to detect the presence of a target. If the tasks cannot be executed together as well as each can be done alone, then the two tasks must compete for the same resources. The pattern of interference can provide a measure of the amount of attentional resources expended in such situations (Kerr, 1973; Norman and Bobrow, 1975).

Two paradigms have been used to study individual differences in the deployment of attention; both depend on some form of computer control for measuring reaction time and errors on a real-time basis. In the "secondary task" paradigm people are given two tasks singly, and then asked to perform them together. In "switching" or "filtering" paradigms people must either switch from monitoring one source of input to monitoring another (e.g. switch from monitoring signals in one ear to monitoring signals presented to the other) or they must monitor one signal source while ignoring another. The presence of a secondary task may introduce a source of individual variation that is not apparent when people perform tasks singly. This could be either because the increased demands of the secondary task provide an accurate measure of the total attentional resources available to a person (Hunt and Lansman, 1982) or because there are reliable individual differences in the ability to allocate attention to competing tasks (Stankov, 1983).

A somewhat different view of attention stresses the control a person has over attention allocation, rather than the total amount of attentional resources available. Gopher (1982) has developed this idea extensively using a dichotic listening paradigm. The examinee listens to the dichotically presented streams of signals. On cue s/he must switch rapidly from attending to the information presented in one ear to attending to information presented in the other. The dependent variable is the speed with which the switch of attention can be executed. Several studies have shown that the attention switching paradigm can be used to discriminate good and bad operators of a variety of machines, ranging from buses to airplanes. The correlations are not large, and the validity of prediction may depend upon the examinee's familiarity with the criterion task. On the other hand, the classification decisions being made in these situations may involve substantial economic commitment. (This is certainly the case in pilot training.) Therefore even small increases in predictability can be valuable.

Given the data now in hand, measures of attentional resources and control warrant serious consideration for inclusion in some classification tests, especially if the goal is to predict performance in situations where rapid decision making is required. The way that a test should be used, though, will depend very much upon the particular situation, because there are a number of theoretical and practical questions to be resolved. Perhaps the most important of these is the issue of situational generality of attention measures. Consider Gopher's procedure. The paradigm involves auditory perception, but is presented as a test of the ability to control attention in general. A recent study by Lansman, Poltrock, and Hunt (1983) indirectly questions this assumption. College students took a variety of auditory and visual tests that required the control of attention. There was evidence for separate factors for control of auditory and visual attention, rather than for a general "attentional control" factor. While Gopher's test was not among those used by Lansman et al., their results do suggest that further studies should be done to determine the factorial complexity of 'attention' itself.

Attention is said to 'wax and wane', thus implying that it is subject to day to day, and perhaps minute to minute variability. While variability might be of interest in itself, (See the remarks above concerning reaction time), any test to test variability poses administrative problems, because extended testing is required. Design problems are also posed because the investment of attention is no more directly observable than is the investment of intelligence. Attention is evaluated by performance on a task. Individual differences in skill on the task are confounded with individual differences on the amount of attention devoted to it. This very much complicates the designs required to isolate measures of attention from measures of task specific skills (Ackerman, Wickens, Schneider, 1982; Hunt and Lansman, 1982). Design issues, rather than technology, may prove the limiting feature in efforts to evaluate attentional capabilities. Nevertheless, given the data now in hand, we regard the extension of testing to the evaluation of individual differences in attention as one of the most promising applications of computer controlled testing.

The Criterion Issue

Validation is one of the thorniest problems in assessment. What psychologists would like to do is to talk about the relation between test scores and "performance" in some extra-laboratory task.

Establishing extra-laboratory performance is usually a substantial problem. Very few aspects of everyday life are evaluated in anything like an objective manner. Perhaps the easiest situation to deal with is the prediction of academic performance, where grades are regularly assigned. Yet anyone only mildly familiar with academics is uncomfortable with grade point averages as an evaluation of student performance. The situation is even worse in the workplace, where competence must often be measured by such surrogates as economic reward or observer ratings, variables that are subject to influences other than personal ability. From a narrow point of view, the validation problem for computer presented tests is no different than the validation problem for any other test. The problem lies in the validation score, not the test score.

Computers could be used to collect measures of human behavior in extra-laboratory situations. The most obvious, and perhaps the most controversial, use is to monitor individual work performance. This is feasible in industries where computer readable records are already used for management control of inventories, customer flow, or other resources. Examples are check-out counters in supermarkets, air traffic control centers, and airplane ticket sales. The recording devices used to keep track of goods or paperwork could be used to estimate human productivity. Such an application raises serious social issues, which we shall not comment upon. We do point out that, seen strictly from the viewpoint of a psychologist interested in assessment, keeping records on individual workers in their normal work environment is an attractive solution to some of the problems that have plagued psychometricians for years. Because the measures would be directly related to work, they would be face valid. There is a danger, of course, that measures that could be recorded would assume a "psychological validity" that they ought not have, to the expense of other, less easily recorded aspects of performance, but this argument could be made against the collection of any less-than-perfect performance measure. A second feature of continuous monitoring is that it records normal performance. Overt testing, on the other hand, is likely to yield measures of (near) maximal performance. The prediction of average performance may be much more important in the workplace. In the past, little work has been done on this because somewhat paradoxically. average performance is more difficult to assess in an objective manner than is peak performance.

In training situations many of the advantages of continuous assessment can be retained without incurring some of the social disadvantages. This is particularly true if the training is to be conducted using simulation techniques that are themselves under computer control, since it may be easier to include computer monitoring of performance in a simulated task than in an actual task.

The combination of assessment with computer controlled simulation may represent an important extension of testing to the workplace. Simulations are becoming increasingly popular training devices, especially in situations where training on the criterion task itself is not feasible. Furthermore, simulation training sessions are less artificial than test situations. A test must be relatively brief and must be meaningful for untrained people. A simulation task may take hours, and it may be practical to train people for several days before they can operate the simulator. Simulator performance itself must be evaluated against actual performance. Once this is done, though, the simulator stands at an intermediate point between brief, inexpensive testing and the expensive, error prone evaluation of extra-laboratory performance. It may be economically feasible to screen candidate tests by examining the correlation between test and simulator performance. Tests that pass this examination can then be evaluated by the more expensive process of collecting data on test and extra-laboratory work performance.

Concluding Remarks

If the decision to move to computer controlled testing is to be made either on the grounds of ease of administration or greater accuracy of evaluation of what we now evaluate, then the decision should be dictated by economic considerations. Given that there exist two tests of the same psychological function,one computerized and one in a more traditional format, which is the most cost effective? Psychological issues are raised only when it is asserted that computerized tests are 'better' measures of cognition than are conventionally administered tests. This is a claim for a psychological gain that is distinct from economic gain.

Providing that the purpose of testing is for personnel screening, it does not appear that there is very much psychological gain to be obtained from computer controlled testing of reasoning, verbal intelligence, or psychomotor abilities. The psychological functions that make up these domains are either not changed by the format of the test, or (in the case of psychomotor functioning) are highly specific to the particular task being used. We are somewhat more optimistic about the use of computer controlled testing if the purpose of an examination is diagnosis of individual performance. In this case much more extensive testing is possible, and computerized administration may make it possible to tap aspects of a person's language processing or problem solving behavior that would not be reliably revealed over a few minutes testing, regardless of the format used.

Computer controlled testing is likely to have much more of an impact on the study of individual differences in spatial-visual reasoning, memory, and on the study of individual differences in attention. In these three fields there are a number of interesting test situations that are simply not feasible without the use of computer controlled testing stations. Research on the development and validation of such test procedures could begin today.

Perhaps one of the most challenging questions for future research is whether or not learning potentials can be evaluated by having people solve problems while being coached by a computer program. At present, questions about learning potential are asked almost exclusively about children, and the research presumes an interaction between the child and a dedicated, highly skilled instructor. It is conceivable, but by no means certain, that interactive computing techniques could be used to identify learning potential in adults. It is unlikely that research in this area would result in procedures that would be useful for initial personnel screening, because the time required to assess learning potential is likely to be on the order of hours. It may be possible to assess adult learning potential in diagnostic situations, where the organization has a commitment to the individual. Given the anticipated reduced size of the workforce over the next fifty years, combined with the rapid introduction of new technology, diagnostic testing may become much more widespread than it is today. If so, high priority should be given to research on the assessment of adult learning potential.

Computers are fascinating, flexible, and highly useful tools. In presenting a somewhat mixed evaluation, we feel almost as if we are taking a stand against progress. Nevertheless, we feel that caution is in order. Computers do raise many possibilities for testing and training. Whether these possibilities are useful ones, though, is not a question that can be answered by demonstrating engineering feasability. The questions that have to be answered are psychological, and the answers can only be obtained by appropriate psychological research.

ff

Author Notes

The preparation of this paper was supported by the Office of Naval Research, Contract N00014-80-C-0631 to the University of Washington (Earl Hunt, Principal Investigator) and by Contract N00014-81-C-0616 to the University of California, Santa Barbara (James Pellegrino, Principal Investigator). The preparation of this report was both inspired and greatly facilitated by the discussion at a conference on the use of computers in intelligence testing, sponsored by the Office of Naval Research and held at the University of California, Santa Barbara in March, 1983. We gratefully acknowledge the comments of the participants at that conference and of Dr.

Marshall Farr, of ONR Personnel and Training Program who had originally proposed that these issues should be explored in this and other forums. The report itself, however, is not in any way a summary of the proceedings of the conference, nor would its conclusions necessarily be endorsed by any conference participants other than ourselves. The opinions expressed here are our own, and do not represent the views of the Office of Naval Research or the U.S. Government.

Footnotes

1. One can imagine some extremely exotic computer controlled tests. We restrict our discussion to the sort of testing that would be possible on computer system that one could reasonably expect to be available for very large testing programs. It will be assumed that the terminal consists of a standard 512 x 512 dot, black and white cathode ray tube (CRT) monitor and that the response device is either a typewriter style keyboard, some form of panel operated by button presses or, when specifically mentioned, a light pen or similar analog

tracking device whose position can be sensed by the computer. This definition covers such devices as the paddles and steering wheels associated with computer games, touch sensitive screens, and drawing devices such as the electronic "mouse". We will consider systems with voice generated output, but not systems with speech recognition capability. Finally. no consideration will be given to devices that can recode electrophysiological signals into computer readable form. While active research on the relation between electrophysiological and behavioral measures on cognition is being conducted, we do not believe that electrophysiologicl measures will be utilized in personnel evaluation in the near future. There will, no doubt, be important special exceptions to this statement.

2. Nickerson, R. Comments on papers in the symposium on individual differences. American Educational Research Association Meetings, Montreal, March 1983.

3. Gopher, D. Remarks on individual differences in attention. Presented at the Conference on Individual Differences. Univ. of CA, Santa Barbara. March 1983.

4. Some illusions may be very difficult to produce using the sort of computer testing configuration that we envisage. Virtually all _standard_ display systems presenting computer testing or

communicative use some variety of the standard "television" display, i.e. 520 x 520 dots, refreshed at 60 Hz. Display changes fast enough to produce or vary an illusion may not be possible in all cases.

5. Payne, D. USAF presentation at the Tri-Services Cognitive Sciences Synthesis Conference, Naval Post Graduate School, Monterey, CA, March 1984. Kyllonen, P., Presentation at Office of Naval Research Contractors' Meeting on Individual Differences and Information Processing, Yale University, March, 1984.

6. Schneider, W. Remarks on individual differences in automatic responding. Conference on Individual Differences. Univ. of CA, Santa Barbara, March, 1983.

7. Campione, J. Child Development and Mental Retardation Center Colloquium. University of Washington, February 1984.

ff.

References

Ackerman, P. L., Schneider, W. and Wickens, C. Individual differences and time-sharing ability: A critical review and analysis. Human Attention Research Laboratory Report.

Baddeley, A. D.(1976) The psychology of memory. New York: Basic Books.

Allen, M. J. and Yew, W. (1979) Introduction to Measurement Theory. Monterey, CA: Brooks-Cole.

Boring E. G. "Intelligence as the tests test it" The New Republic, 1923, 34, 35-36.

Brand, C. General intelligence and mental speed: Their relationship and development. In Freidman, M., Das, J. P., and O'Connor, N. (eds.) Intelligence and Learning, New York: Plenum Press, 1981, 589-594.

Bransford, J., Stein, B. S., Vye, N. J., Franks, J. J., Auble, P. A., Mezynski, K. J. and Perfetto, G. A. Differences in approaches to learning: An overview. J. Experimental Psychology: General 1982, 111, 390-398.

Brown, J. S. and Burton, R. B. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Buros, O. K. (1975) Mental measurements yearbook and tests in print II. Highland Park, N. J. : Gryphon Press.

Calvert, E. J. and Waterfall, R. C. A comparison of conventional and automated administration of Raven's Standard Progressive Matrices. International J. of Man-Machine Studies, 1982, 17, 305-310.

Campione, J., Brown, A. L. and Ferrara. R. A. Intelligence and mental retardation. In Sternberg, R. J. Handbook of Human Intelligence, Cambridge: Cambridge U. Press, 1982.

Carpenter. P. A. and Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 1975, 82, 45-73.

Carr, A. C., Wilson, S. L, Ghorb, A., Angell, R. J., and Mums, R. T., Automated testing of geriatric patients using a microcomputer based system. (1983) International journal of man machine systems , 17, 297-300.

Carroll. J. B.(1971) Development of native language skills beyond the early years in Reed, C. B. (ed.) The learning of language . New York: Appleton-Century-Crofts.

Carroll. J. B. (1979) Psychometric approaches to the study of language abilities in Fillmore, C. J., Kemper, P., and Wang, S. Y. Individual differences in language ability and language behavior . New York: Academic Press.

Carroll. J. B. The measurement of intelligence. In Sternberg, R. J. Handbook of Human Intelligence, Cambridge: Cambridge U. Press, 1982.

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

Chase, G. and Eriksson, K. A. Skilled memory in cognitive skills and their acquisition. Anderson J. (ed.). Hillsdale, New Jersey: Lawrence Erlbaum Association, 1981.

Coren, S. and Girgus, J. S. (1978) Seeing is deceiving. The psychology of visual illusion. Hillsdale, N.J.: L. Erlbaum Associates.

Dillon R. F. and Stevenson-Hicks, R. Effects of item difficulty and method of test administration on eye scan patterns during analogical reasoning. Technical report #1, Contract N66001-80-C0467. Southern Illionois Univ., 1981.

Dixon, P. D. Plans and written directions for complex tasks. J. Verbal Learning and Verbal Behavior, 1982, 21, 70-84.

Dunlap, W. P., Bilodeau, I., and Jones, M. B. Factors appearing late in practice. Organizational Behavior and Human Performance, in press.

Egan, D. Testing based on understanding: Implications from studies of spatial ability. Intelligence, 1979, 3, 1-15.

Egan, D. E. and Grimes-Farrow, D. D. (1982) Differences in mental representations spontaneously adopted for reasoning. Memory and cognition, 10, 297-307.

Estes, W. K. Learning, memory and intelligence. In Sternberg, R. J. (ed.) Handbook of Human Intelligence. Cambridge: Cambridge U. Press, 1982.

Feurstein, R. Instrumental enrichment: An intervention program for cognitive modifiability. Baltimore, MD: University Park Press, 1980.

Fillmore, C. J. On fluency. In Fillmore, C. J., Kempler, D., and

Wang, W. S-Y (eds.) Individual Differences in Language Ability and Language Behavior. New York: Academic Press, 1979.

Fisk, A. D. and Schneider, W. Category and word search: Generalizing search principles to complex processing. J. Experimental Psychology: Learning, Memory, and Cognition, 1983, 9, 177-195.

Fleishman, E. A. Individual differences in motor learning. In Gagne, P. M. (ed.) Learning and Individual Differences. Columbus, OH: Carles G. Merrill. 1967.

Frederiksen, J. R. A componential theory of reading skills and their interactions. In Sternberg, R. (ed.) Advances in the psychology of human intelligence. Hillsdale, N. J. L. Erlbaum Associates. 125-180.

Goldman, S. R., and Pellegrino, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In kR. J. Sternbery (Ed.), Advances in the psychology of human intelligence. vol.2. Hillsdale, NJ: Erlbaum.

Gopher, D. A. A selective attention test as a predictor of success in flight training. Human Factors, 1982, 24, 173-183.

Green B. F. A primer of testing. American Psychologist, 1981, 36, 1001-1011.

Green B. F., Bock, R. D., Humphreys, L. G., Lin, R. L. and Reckase, M. D., Evaluation plan for the computerized adaptive vocational aptitude battery. Research Report 82-1, Contract N00014-80-K-304, Office of Naval Research, 1982.

Gross, M. L. The Psychological Society, New York: Random House, 1978.

Hirst, W. The amnesic syndrome: Descriptions and explanations. Psychological Bulletin, 1982, 91, 435-460.

Holzman, T., Pellegrino, J. W., and Glaser, R. (1983). Cognitive variables in series completion. Journal of Educational Psychology, 75, 602-617.

Horn, J. L. and Donaldson, G. Cognitive Development II: Adult development of human abilities. In Kagan, J. and Brim, O. G. Consistency and change in human development, Cambridge, Mass.: Harvard U. Press, 1980.

Hunt, E. What kind of computer is man? Cognitive Psychology. 1971, 2, 57-98.

Hunt, E. Mechanics of Verbal Ability, Psychological Review, 1978, vol. 85. 2, 109-130.

Hunt, E. Intelligence as an information processing concept. British J. Psychology, 1980, 71, 449-474.

Hunt, E. The design of a robot mind: a theoretical approach to issues in intelligence. In M. Friedman, J. R. Das, and N. O'Connor (eds.) Intelligence and Learning, 1981, 14, 459-478.

Hunt, E. Towards new ways of assessing intelligence. Intelligence, 1982, 6, 231-240.

Hunt, E. Verbal Ability in Sternberg, R. J. (ed.) Human abilities: An information processing approach, in press.

Hunt, E., Davidson, J. and Lansman, M. Individual differences in long-term memory access. Memory and Cognition, 1981, 9, (6), 599-608.

Hunt, E. and Hertzog, C. Age related changes in cognition during the working years. U. of WA Psychology Dept. Technical Report, 1981.

Hunt, E. and Lansman, M. Individual differences in attention. In Sternberg, R. J. (ed.) Psychology of Human Intelligence. Hillsdale, NJ: L. Erlbaum Assoc., 1982.

Jensen, A. R. (1980) Bias in mental testing. Glencoe, Ill: Free Press.

Jensen, A. R. and Munro, E. Reaction time, movement time, and intelligence. Intelligence, 1979, 3, 121-126.

Johnson-Laird, P. N.(1983) Mental models . Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. and Steedman, M. The psychology of syllogisms. Cognitive Psychology, 1978, 10, 64-99.

Kahneman, P. Attention and Effort. Englewood Cliffs, NJ: Prentice Hall, 1973.

Kennedy, R. S., Carter, R. C. and Bittner, A. C. A catalogue of performance evaluation tests for environmental research. Proceedings 24th Annual Meeting of the Human Factors Society, Los Angeles, CA, Oct. 1980.

Kerr, B. Processing demands during mental operations. Memory and Cognition, 1973, 1, 401-412.

Klapp, S. T., Marshburn, E. A. and Lester P. T. Short term memory does not involve the "working memory" of information processing: The decline of a common assumption. J. Experimental Psychology: General, 1983, 112. 240-264.

Lansman, M., Donalson, G., Hunt, E. and Yantis, S. Ability and factors and cognitive processes. Intelligence, 1982, 6, 347-386.

Lansman, M., Poltrock, S. E., and Hunt, E. Individual differences in the ability to focus and divide attention. Intelligence, in press.

Loftus, G. and Loftus, E. F. (1983) The mind at play . New York: Basic Books.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

MacLeod, C. M., Hunt, E., and Mathews, N. N. Individual differences in the verification of sentence-picture relationships. J. Verbal Learning and Verbal Behavior, 1978, 17, 493-507

Marshall-Mies, J. Yarkin-Levin, K. and Schemmer, F. Development of computerized tests for selection of system operations. Paper given at American Psychological Association meeting, Anaheim, CA, August, 1983.

Matarazzo, J. D. Wechsler's measurement and appraisal of adult intelligence (5th edition) Baltimore, MD: Williams and Wilkins, 1972.

Matarazzo, J. D. Computerized Psychological Testing. Science, 1983, 1.

McGee, M. G. Human Spatial Abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 1979, 86, 889-918.

Mumaw, R. J., and Pellegrino, J. W. (in press). Individual differences in complex spatial processing. Journal of Educational Psychology.

Navon, D. and Gopher, D. On the economy of the human processing system. Psychological Review, 1979, 86, 214-255.

Nettlebeck, T. and Kirby, N. H. (1983) Measures of timed performance and intelligence. Intelligence 7, 39-52.

Newell. A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. P. (ed.) Cognitive skills ad their Acquisition. Hillsdale, NJ: L. Erlbaum Associates, 1981.

Norman, D. and Bobrum, D. On data limited and resource limited processes. Cognitive Psychology, 1975, 7, 44-64

Pellegrino, J. W., and Kail, R. Jr. Process analyses of spatial aptitude. In Sternberg, R. J. (ed.) Advances in the Psychology of Human Intelligence, vol. 1. Hillsdale, NJ: L. Erlbaum and Assoc., 1982.

Poltrock, S. E., Lansman, M. and Hunt, E. Automatic and controlled processes in auditory target detection. J. Experimental Psychology: Human Perception and Performance, 1982, 8, 37-45.

Posner, M. Z. (1978) Chronometric explorations of the mind . Hillsdale, N. J. L. Erlbaum Associates.

Potter, M. C., Krull. J. F. and Harris, C. (1980) Comprehension and memory in rapid sequential reading. In Nickerson, R. S. (ed.) Attention and performance

VIII. Hillsdale, N. J., L. Erlbaum Associates.

Rips, L. J. Cognitive processes in propositional reading. Psychological Review, 1983, 90, 38-71.

Sarason, I, (ed.) (1980) Test anxiety: theory, research, and applications. Hillsdale, N. J., L. Erlbaum Associates.

Schneider, W., and Shiffrin R. M. Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 1977, 84. 1-66.

Scribner, S. Modes of thinking and ways of speaking: Culture and logic reconsidered. In Johnson-Laird, S. and Wason, P. C. (eds.) Thinking: Readings in Cognitive Science. Cambridge: Cambridge U. Press, 1977.

Shepard, R. and Cooper, L. Mental images and their transformations. Cambridge, MA: MIT Press, 1983.

Siegler, R. S. and Richards, D. D. (1982) The development of intelligence in Sternberg, S. J. (ed.) Handbook of human intelligence. Cambridge, England: Cambridge University Press, 897-973.

Sleeman, D. H. Inferring student models for intelligent computer aided instruction. In Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (eds.) Machine Learning: An Artificial Intelligence Approach. Palo Alto, CA: Tioga Press, 1983, 483-510.

Spearman, C. The abilities of man. New York: MacMillan, 1927.

Squire, L. R. Comparisons between forms of amnesia: Some deficits are unique to Korsakoff's syndrome. J. Expt'l Psychology: Learning, Memory, and Cognition, 1982, 8, 560-571.

Stankov, L. The role of competition in human abilities revealed through auditory tests. Multivariate behavioral research monographs, 1983.

Sternberg, R. J. Intelligence, information processing and analogical reasoning. Hillsdale, NJ: L. Erlbaum Assoc, 1977.

Sternberg, R. J. Sketch of a componential subtheory of human intelligence. The Behavioral and Brain Sciences, 1980, 3, 573-584.

Sternberg, R. J. Intelligence and Nonentrenchment. J. Educational Psychology, 1981, 12, 10-17.

Sternberg, R. J. Reasoning, problem solving, and intelligence. In Sternberg, R. J. (ed.) Handbook of Human Intelligence. Cambridge, England: Cambridge U. Press, 1982.

Sternberg, R. J., Conway, B. E., Ketron, J. L. and Bernstein, M. People's conceptions of intelligence. J. of Personality and Social Psychology: Attitudes and Social Cognition, 1981, 41, 37-55.

Sternberg, R. J. and Gardner, M. K. Unities in inductive reasoning. J. Expt'l Psychology: General, 1983, 112, 80-116.

Sunderland, A., Harris, J. E. and Baddeley A. D. Do laboratory tests predict everyday memory? A neuropsychological study. J. Verbal Learning and Verbal Behavior, 1983, 22, 341-357.

Thorndyke, P. W. and Goldin, P. W. Ability differences and Cognitive mapping skill. Rand Corporation, Santa Monica, CA Technical report N-1664-ARMY, March 1981.

Underwood, B. J., Boruch, R. F. and Malmi, R. D. Composition of episodic memory. J. Expt'l Psychology: General, 1978, 107, 393-419.

Vandenberg, S. G. and Kuse, A. R. Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and motor skills, 1978, 47, 599-604

Vellution, F. R. (1979) Dyslexia, Theory and research . Cambridge, MA: M.I.T. Press.

Vernon, P. A. Speed of information processing and general intelligence. Intelligence 7, 53-70.

Vygotsky, L. S. Mind in society: The development of higher psychological processes. (M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, Eds. and translators) Cambridge, MA: Harvard U. Press, 1978..

Watts, K., Baddeley, A. and Williams, M. Automated tailored testing using Raven's matrices and the Mini Hill Vocabulary Tests: A comparison with manual administration. International J. of Man-Machine Studies, 1982, 17, 331-344.

Whitely, S. E. Measuring aptitude processes with multi component latent trait models. J. Educational measurement, 1981, 18, 67-84.

Wickens, C. D. The structure of attentional resources. In Nickerson,

R. S. (ed.) Attention and Performance VIII. Hillsdale, NJ: Lawrence Erlbaum Associates., 1979.

Witkem, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., and Wapner, S. (1954) Personality through perception New York: Harper.

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152

1 Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Lt. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32508

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Dr. Tom Duffy
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. Jim Hollan
Code 14
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

Navy

1 Dr. Peter Kincaid
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508

1 Dr. Joe McLachlan
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Office of Naval Research
Code 433
800 N. Quincy SStreet
Arlington, VA 22217

1 Office of Naval Research
Code 441NP
800 N. Quincy Street
Arlington, VA 22217

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
OP 115
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

Navy

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode, Director
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

Army

1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Marshall Narva
US Army Research Institute for the
Behavioral & Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commander, U.S. Army Research Institute
for the Behavioral & Social Sciences
ATTN: PERI-BR (Dr. Judith Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 20333

1 Joseph Psotka, Ph.D.
ATTN: PERI-1C
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Wisher
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Bryan Dallman
AFHRL/LRT
Lowry AFB, CO 80230

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Joseph Yasatuke
AFHRL/LRT
Lowry AFB, CO 80230

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Jordan Grafman, Ph.D.
Department of Clinical Investigation
Walter Reed Army Medical Center
6825 Georgia Ave., N. W.
Washington, DC 20307

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defens for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civilian Agencies

1 Dr. Patricia A. Butler
NIE-BRN Bldg, Stop # 7
1200 19th St., NW
Washington, DC 20208

1 Dr. Paul G. Chapin
Linguistics Program
National Science Foundation
Washington, DC 20550

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Edward Esty
Department of Education, OERI
MS 40
1200 19th St., NW
Washington, DC 20208

1 Edward J. Fuentes
National Institute of Education
1200 19th Street, N. W.
Washington, DC 20208

1 Gloria Gilmer
National Institute of Education
1200 19th St. N.W.
Mail Stop 7
Washington, DC 20208

1 TA&E, T&L
National Institute of Education
1200 19th St., NW
Washington, DC 20208

1 Dr. Arthur Melmed
724 Brown
U. S. Dept. of Education
Washington, DC 20208

1 Dr. Andrew R. Molnar
Office of Scientific and Engineering Personnel and Education
National Science Foundation
Washington, DC 20550

1 Dr. Ramsay W. Selden
National Institute of Education
1200 19th St., NW
Washington, DC 20208

Civilian Agencies

1 Chief, Psychological Research Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. John Black
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Glenn Bryan
6208 Poe Road
Bethesda, MD 20817

1 Dr. Jaime Carbonell
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Private Sector

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94306

1 Dr. Michael Cole
University of California
at San Diego
Laboratory of Comparative
Human Cognition - D003A
La Jolla, CA 92093

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Mr. Wallace Feurzeig
Department of Educational Technology
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 02238

1 Dr. Dexter Fletcher
WICAT Research Institute
1875 S. State St.
Orem, UT 22333

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Don Gentner
Center for Human Information Processing
University of California, San Diego
La Jolla, CA 92093

1 Dr. Dedre Gentner
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 02138

Private Sector

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Josph Goguen
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

1 Dr. Daniel Gopher
Faculty of Industrial Engineering
& Management
TECHNION
Haifa 32000
ISRAEL

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Barbara Hayes-Roth
Department of Computer Science
Stanford University
Stanford, CA 95305

1 Glenda Greenwald, Ed.
Human Intelligence Newsletter
P. O. Box 1163
Birmingham, MI 48012

1 Dr. Lloyd Humphreys
Department of Psychology
Universi of Illinois
603 East Daniel Street
Champaign, IL 61820

Private Sector

1 Dr. Marcel Just
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Scott Kelso
Haskins Laboratories, Inc
270 Crown Street
New Haven, CT 06510

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Stephen Kosslyn
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138

1 Dr. Pat Langley
The Robotics Institute
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Marcy Lansman
The L. L. Thurstone Psychometric
Laboratory
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Jim Levin
University of California
at San Diego
Laboratory fof Comparative
Human Cognition - D003A
La Jolla, CA 92093

Private Sector

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Don Lyon
AFHRL/OT (UDRI)
Williams AFB, AZ 85225

1 Dr. Jay McClelland
Department of Psychology
MIT
Cambridge, MA 02139

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. Donald A Norman
Cognitive Science, C-015
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara, CA 93106

Private Sector

1 Dr. Nancy Pennington
University of Chicago
Graduate School of Business
1101 E. 58th St.
Chicago, IL 60637

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309

1 Dr. Mike Posner
Department of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Fred Reif
Physics Department
University of California
Berkeley, CA 94720

1 Dr. Lauren Resnick
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 1521

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
Murray Hill, NJ 07974

1 Dr. William B. Rouse
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Michael J. Samet
Perceptronics, Inc
6271 Variel Avenue
Woodland Hills, CA 91364

Private Sector

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

1 Dr. Alan Schoenfeld
Mathematics and Education
The University of Rochester
Rochester, NY 14627

1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Kathryn T. Spoehr
Psychology Department
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

1 Dr. Perry W. Thorndyke
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

Private Sector

1 Dr. Kurt Van Lehn
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 Dr. Keith T. Wescourt
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025

1 William B. Whitten
Bell Laboratories
2D-610
Holmdel, NJ 07733

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Thomas Wickens
Department of Psychology
Franz Hall
University of California
405 Hilgarde Avenue
Los Angeles, CA 90024

1 Dr. Mike Williams
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 94304