DOCUMENT RESUME

ED 327 547 TM 015 371

AUTHOR Kyllonen, Patrick C.; Christal, Raymond E.
TITLE Cognitive Modeling of Learning Abilities: A Status Report of LAMP.
INSTITUTION Air Force Human Resources Lab., Brooks AFB, Tex. Manpower and Personnel Div.
REPORT NO AFHRL-TP-87-66
PUB DATE Feb 88
NOTE 47p.; This report in an Interim Technical Paper for the period November 1986 to November 1987.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability Identification; Achievement Tests; Computer Assisted Testing; Expert Systems; *Federal Programs; Individual Differences; Knowledge Level; *Learning Processes; Memory; *Military Personnel; Models; Personnel Selection; Problem Solving; *Research Projects; Skill Analysis; *Test Validity
IDENTIFIERS *Learning Abilities Measurement Program

ABSTRACT

Research activities underway as part of the Air Force's Learning Abilities Measurement Program (LAMP) are described. A major objective of the program is to devise new models of the nature and organization of human abilities, that could be applied to improve personnel selection and classification systems. The activities of the project have been divided into two categories. The first category is concerned with identifying fundamental learning abilities by determining how learners differ in their abilities to think, remember, solve problems, and acquire knowledge and skills. A four-source framework is adopted which assumes that observed learner differences are due to differences in processing speed; processing capacity; and the breadth, extent, and accessibility of conceptual knowledge and procedural skills. The second category of research activities is concerned with validating new models of learning abilities. To do this, several computerized intelligent tutoring systems have been built that serve as mini-courses in technical areas, such as computer programming and troubleshooting electrical circuits. A major objective of this part of the program is to develop principles for producing indicators of student learning progress and achievement. These indicators will serve as learning outcome measures for ability test validation. Four figures and a 76-item list of references are included. (Author/TJH)

AFHRL-TP-87-66

AIR FORCE

HUMAN

RESOURCES

COGNITIVE MODELING OF LEARNING ABILITIES: A STATUS REPORT OF LAMP

Patrick C. Kyllonen

Institute for Behavioral Research
University of Georgia
Athens, Georgia 30602

Raymond E. Christal

Universal Energy Systems
4401 Dayton-Xenia Road
Dayton, Ohio 45432-1894

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

February 1988
Interim Technical Paper for Period November 1986 - November 1987

Approved for public release; distribution is unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

DAN WOLTZ
Contract Monitor

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

DANIEL L. LEIGHTON, Colonel, USAF
Chief, Manpower and Personnel Division

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION
Unclassified

1b RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFHRL-TP-87-66

6a. NAME OF PERFORMING ORGANIZATION
Institute for Behavioral Research

6b OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Manpower and Personnel Division

6c. ADDRESS *(City, State, and ZIP Code)*
University of Georgia
Athens, Georgia 30602

7b. ADDRESS *(City, State, and ZIP Code)*
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
Air Force Human Resources Laboratory

8b OFFICE SYMBOL *(If applicable)*
HQ AFHRL

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F41689-84-D-0002/58420360
S-744-031-001 & S-744-049-001

8c. ADDRESS *(City, State, and ZIP Code)*
Brooks Air Force Base, Texas 78235-5601

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
61102F	2313	T1	33

11. TITLE *(Include Security Classification)*
Cognitive Modeling of Learning Abilities: A Status Report of LAMP

12 PERSONAL AUTHOR(S)
Kyllonen, P.C.; Christal, R.E.

13a. TYPE OF REPORT
Interim

13b. TIME COVERED
FROM Nov 86 TO Nov 87

14. DATE OF REPORT *(Year, Month, Day)*
February 1988

15. PAGE COUNT
44

16. SUPPLEMENTARY NOTATION
To appear in R. Dillon & J.W. Pellegrino (Eds.), Testing: Theoretical and Applies Issues. San Francisco: Freeman.

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	10	
05	09	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
cognition
cognitive ability
computerized testing
individual differences
learning
learning ability
learning abilities measurement program (LAMP)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This paper describes some of the research activities underway as part of the Air Force's Learning Abilities Measurement Program (LAMP). A major objective of this basic research project is to devise new models of the nature and organization of human abilities with the long-term goal of applying these models to improve current personnel selection and classification systems. The activities of the project have been divided into two categories. The first category is concerned with identifying fundamental learning abilities by determining how learners differ in their abilities to think, remember, solve problems, and acquire knowledge and skills. From research already completed, a four-source framework has been established that assumes observed learner differences to be due to differences in processing speed; processing capacity; and the breadth, extent, and accessibility of conceptual knowledge and procedural skills. The second category of research activities is concerned with validating new models of learning abilities. To do this, a number of computerized intelligent tutoring systems have been built that serve as mini-courses in technical areas such as computer programming and troubleshooting electrical circuits. A major objective of this part of the program is to develop principles for producing

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Nancy J. Allin, Chief, STINFO Office

22b. TELEPHONE *(Include Area Code)*
(512) 536-3877

22c OFFICE SYMBOL
AFHRL/TSR

DD Form 1473, JUN 86 *Previous editions are obsolete*

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. (Concluded)

indicators of student learning progress and achievement. These indicators will serve as the learning outcome measures against which newly developed learning abilities tests will be evaluated in future validation studies.

AFHRL Technical Paper 87-66

February 1988

COGNITIVE MODELING OF LEARNING ABILITIES:
A STATUS REPORT OF LAMP

Patrick C. Kyllonen

Institute for Behavioral Research
University of Georgia
Athens, Georgia 30602

Raymond E. Christal

Universal Energy Systems
4401 Dayton-Xenia Road
Dayton, Ohio 45432-1894

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed by
William C. Tirre
Learning Abilities Research Function

Submitted for publication by

William R. Ercoline, Lieutenant Colonel, USAF
Chief, Cognitive Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

This paper outlines some of the research activities underway as part of the Air Force's Learning Abilities Measurement Program (LAMP). The major goal of the project is to devise new models of the nature and organization of human abilities with the long-term goal of applying those models to improve current personnel selection and classification systems. As an approach to this ambitious undertaking, we have divided the activities of the project into two categories. The first category is concerned with identifying fundamental learning abilities by determining how learners differ in their abilities to think, remember, solve problems, and acquire knowledge and skills. From research already completed, we have established a four-source framework that assumes that observed learner differences are due to differences in *processing speed*; *processing capacity*; and the breadth, extent, and accessibility of *conceptual knowledge* and *procedural and strategic skills*. The second category of research activities is concerned with validating new models of learning abilities. To do this, we are building a number of computerized intelligent tutoring systems that serve as mini-courses in technical areas such as computer programming and electronics troubleshooting. A major objective of this part of the program is to develop principles for producing indicators of student learning progress and achievement. These indicators will serve as the learning outcome measures against which newly developed learning abilities tests will be evaluated in future validation studies.

PREFACE

Development of this paper was supported by the Air Force Learning Abilities Measurement Program (LAMP), a multi-year program of basic research conducted at the Air Force Human Resources Laboratory (AFHRL) and sponsored by the Air Force Office of Scientific Research. The goals of the program are to specify the basic parameters of learning ability, to develop techniques for the assessment of individuals' knowledge and skill levels, and to explore the feasibility of a model-based system of psychological assessment. Support was provided by AFHRL and the Air Force Office of Scientific Research, through Universal Energy Systems, under Contract No. F41689-84-D-0002/58420360, Subcontract No. S-744-031-001, and Subcontract No S-744-049-001. We thank Valerie Shute, William Tirre, and William Alley of AFHRL for their comments on this paper, and we give a special acknowledgement to Dan Woltz of AFHRL for many long and thorough discussions of the issues addressed herein.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

Considerable headway has been made during the last decade in our understanding of human cognition. This has led to speculation that it is only a matter of time before an improved technology for gauging individuals' intellectual proficiencies will be developed. The stakes are high: Psychological testing of cognitive proficiency is presently widespread in industry, the schools, and the military. Improved tests would have a profound economic impact in cutting education and training costs and enabling a more efficient and fair system of personnel utilization. Although the concept of psychological testing must certainly be considered one of psychology's true success stories, it is also primarily a past accomplishment. Systematic studies of predictive validity have shown that today's aptitude tests are no better than those available shortly after World War II (Christal, 1981; Kyllonen, 1986).

But even if it is agreed that forces are conspiring to usher in a new era of cognitive testing, there still is considerable debate on exactly what form these new cognitive tests will take. On one side of the debate, some argue that what cognitive psychology has to offer is a rationale and a methodology for measuring basic information processing components (Detterman, 1986; Jensen, 1982; Posner & McLeod, 1982). According to this view, the cognitive test battery of the future would consist of measures of speed of retrieval from long-term memory, short-term memory scanning rate, probability of transfer from short- to long-term storage, and the like. On the opposite end of the debate are those who suggest that the fundamental insight of cognitive science is that cognitive skill reflects primarily knowledge rather than general processing capabilities. This perspective has led to calls for testing intermingled with instruction, testing aimed at measuring what students know and what they have learned in the context of their current instructional experience (Embretson, in press; Glaser, 1985). This has been called *steering testing* (Lesgold, Bonar, & Ivill, 1987) or *apprenticeship testing* (Collins, 1986). Between these positions are those who propose new kinds of cognitive tests that are not radically different from existing ones but perhaps richer and more diverse in what they measure (Hunt, 1982; Hunt & Pellegrino, 1984; Sternberg, 1981b).

In this paper, we provide a status report of one ongoing program of research, the Learning Abilities Measurement Program (LAMP), that has been concerned with developing new methods for measuring cognitive abilities. We discuss some of our early thinking on the implications of cognitive psychology for testing, and how we have adjusted our ideas in light of data collected in our cognitive abilities measurement (CAM) laboratory. We conclude with a brief discussion of CLASS, the Complex Learning Assessment Laboratory, the setting in which we intend to validate the new tests.[1]

II. COGNITIVE THEORY AND APTITUDE TESTING

The idea of grounding psychological testing in cognitive theory is not entirely novel. During the 1970s and 1980s, the Air Force Office of Scientific Research (AFOSR) and especially, the Office of Naval Research (ONR) supported a number of basic research projects which had the explanation of individual differences in learning and cognition as a central goal. This research largely concentrated on the analysis of conventional aptitude tests, probably for two reasons. First, analysis of aptitude tests is important in its own right, as an attempt to determine what it is that such tests measure. But, second, and perhaps more importantly, aptitude tests can be viewed as generic surrogates for tasks tapping more complex, slowly developing learning skills. It is difficult and extremely expensive to identify and analyze the information processing components associated with the acquisition of computer programming skill; so goes the argument: It is far cheaper and more efficient to analyze the seemingly more tractable components of some aptitude test, such as an analogies test, that predicts success in computer programming. And the fact that tests do such a good job in predicting training outcomes can be taken as evidence that pretty much the same cognitive components are involved in both test-taking and learning.

---

[1] This paper does not review the research accomplished by William Tirre and Linda Elliott concerning individual differences in text comprehension. Readers interested in this area are referred to Tirre and Elliott (1987).

The wave of aptitude research that was motivated by these considerations did not lead directly to improvements in existing aptitude testing systems, however. A number of new methods and techniques, such as cognitive correlates analysis (Hunt, Frost, & Lunneborg, 1973) and componential analysis (Sternberg, 1977), were developed for analyzing aptitude tests, but the application of these methods did not suggest how the tests themselves might be improved. There have been suggestions that cognitive tasks exported from the experimental psychologist's laboratory might somehow be used to supplement or even replace existing aptitude tests (Carroll, 1981; Hunt, 1982; Hunt & Pellegrino, 1984; Pellegrino & Glaser, 1979; Rose & Fernandez, 1977, Snow, 1979; Sternberg, 1981b), but after almost 10 years, the research still has not been carried out to an extent sufficient for determining whether this is really feasible.

Probably the reason cognitive-based aptitude research has not translated already into better tests is that this has not been a primary goal of the research. Indeed, if the creation of better tests had been the primary goal, the approach of analyzing and decomposing existing tests does not seem very promising. If such research efforts were completely successful, "if the research turned out better than anyone's wildest expectations," at best, new tests would simply duplicate the validity of existing tests.

III. LEARNING ABILITIES MEASUREMENT PROGRAM (LAMP)

In contrast to some of the aptitude research projects previously discussed, our own work in connection with Project LAMP has from its inception been focused on the goal of developing an improved selection and classification system. Our current efforts fall into two categories. First, we are continuing to model basic cognitive learning skills and their interrelationships, and to explore different methods for measuring these skills. Second, we have more recently begun thinking seriously about a system for validating the new cognitive measures. The system involves the extraction of learning indices, both on short-term (1 hour) and long-term (1 week) learning tasks, that will serve as criteria against which the new cognitive measures will be validated. Although we have not yet collected data on the long-term learning tasks, we have set up the laboratory, which consists of 30 computerized tutoring stations. In the remainder of this paper, we discuss these two categories of ongoing LAMP research. We begin with a discussion of studies that have attempted to measure cognitive skills.

Modeling Cognitive Skills: The Four-Source Framework

Much of our work on identifying basic learning skills has centered around what we have called the four-source framework (Kyllonen, 1986). This is the idea that individual differences in a wide variety of learning and performance tasks are due to differences in four underlying sources: (a) effective cognitive *processing speed*; (b) effective *processing capacity*; and the general breadth, accessibility, and pattern of one's (c) conceptual *knowledge* and (d) procedural and strategic *skills*. Figure 1 illustrates these relationships.

We refer to the knowledge and skill components of this model (components [c] and [d]) as *enablers*, in the sense that any learning or performance task can be characterized as consisting of a necessary set of knowledge and skill prerequisites. We refer to the processing speed and working memory components of the model ([a] and [b]) as *mediators*, in the sense that these components mediate the degree to which the learner or problem-solver is able to use his or her knowledge and skills effectively. We have found the four-source framework to be useful in organizing our own as well as others' research and in monitoring our research progress. Further, although we have not yet applied it widely in this fashion, we expect that the system will be useful for task analysis purposes.

Thus far, most of the research we have accomplished in connection with the four-source proposal has been concerned with (a) improving the way in which we measure cognitive skills and (b) determining the dimensionality of the skills and subskills embedded within the four-source model. We now turn to a discussion of the four components, in turn.

Processing Speed

Considerable research on individual differences in cognition over the past 10 years has been concerned with determining the relationship between processing speed and performance on complex

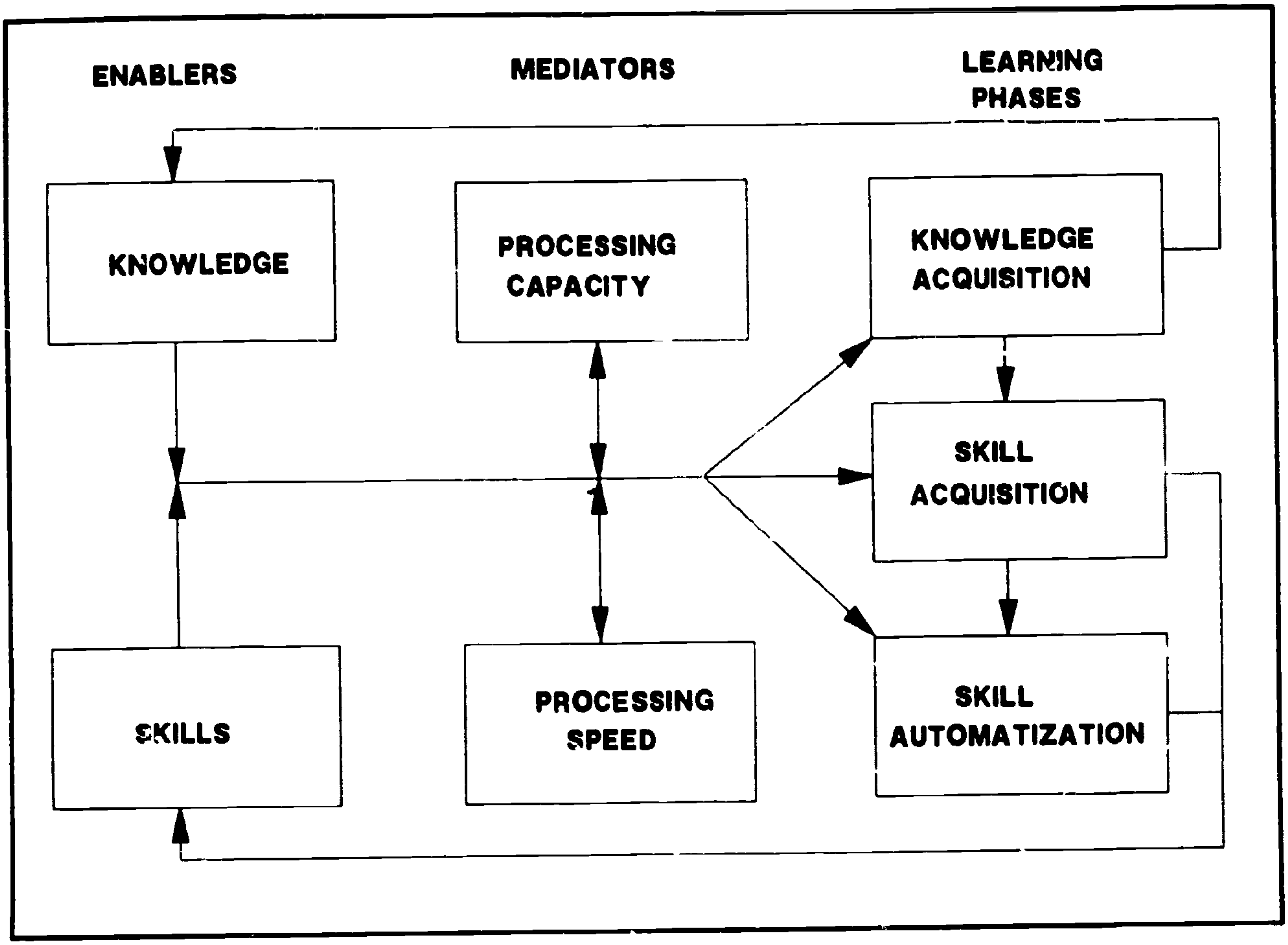


Figure 1. Four-Source Research Framework. Performance in each of the three learning phases (Knowledge Acquisition, Skill Acquisition, and Skill Automatization phases; right side of figure) is presumed to be a function of the *enablers* (Knowledge and Skills), the *mediators* (Processing Capacity and Processing Speed), and whether the prior *learning phase* is complete.

tasks, such as intelligence tests. There are a number of reasons for the high level of interest in processing speed. One is that we now can measure it. The availability of microcomputers as testing instruments makes it feasible to measure, with precision, response time to particular items. Paper-and-pencil tests allowed only gross estimates of response speed. Second, processing speed seems to reflect something basic, something fundamentally a part of all mental activity, and therefore something that might explain the general factor in intelligence, in some sense. Third, since the beginnings of modern cognitive psychology, processing speed has played a major role in cognitive theories in revealing the dynamics of mental processes. Neisser's (1967) book, which is generally considered the kickoff point for the discipline, reported primarily on reaction time studies. Finally, there are operational performance contexts, such as the Air Traffic Controller Workstation or the cockpit, that require efficient processing of considerable data. Understanding the relationship between processing speed and performance in these contexts would have immediate practical payoff.

In our own laboratory, we have conducted a number of studies on processing speed that have focused on both its psychometric properties and its relationship to performance on criterion tasks. Studies have run the gamut in addressing both applied and basic issues. A number of early studies in the project (reported in Kyllonen, 1985) were designed simply to address the question of whether processing speed could be more appropriately characterized as a unitary or multidimensional construct. That is, we addressed the question of whether some people are generally faster information processors than others, or whether it is more appropriate to think in terms of varieties of processing speed. Both positions can be argued for on rational grounds. Much of Jensen's work (Jensen, 1982) at least implicitly presumes a general speed factor. But low correlations between processing speed tasks and measures of general intelligence have led others to propose multiple, correlated processing speed components (e. g., Detterman, 1980).

One way to address the dimensionality question is simply to measure response time on a wide variety of cognitive tests, such as those one finds in the Educational Testing Service (ETS) kit, and perform a factor analysis on the resulting scores. In one study (Kyllonen, Tirre, & Christal, 1985), we did just that and found evidence for both separate reasoning, quantitative, and verbal processing factors, and a higher-order general processing speed factor. Interestingly, we found that although processing speed scores were quite reliable, at least within session, they were not related to accuracy scores on the same tests. Timed versions of the tests thus mix these two separable components of performance in yielding only a single score. There are problems with this approach to testing the dimensionality question, such as how to allow for speed-accuracy trade-off, what to do with response times when the person guessed incorrectly, and so forth But a more substantive problem is that although the findings are suggestive, they fall considerably short of revealing much about the processes that produced them.

Thus, in subsequent work we have restricted our focus (and employed a narrower range of tasks) in the hope of achieving a better process-oriented understanding of the generality question. In these studies, we attempted to identify processing stages, then measure the duration of those stages for individual subjects, then compute the stage inter-correlations. The procedure is best illustrated by example. In the first study (Kyllonen, 1987), we administered a series of tasks that required subjects simply to determine whether two words presented (e.g., *happy-lose*) were similar or dissimilar with respect to valence. *Happy* would be considered a positive-valence word; *lose* would be considered a negative valence word. We presumed that a decision on this task was executed after a series of processing stages. The subject begins by *encoding* one of the words, then encoding the second word. The result of the encoding process is that a symbol representing valence is deposited in working memory for each word. The subject then *compares* those symbols. The result of the comparison process is an implicit assertion that the symbols are either the same or different. A *decision* process then takes the comparison result and translates it into a plan for the execution of the motor response. A *response* process then executes the motor response. Through the method of pre-cueing, which has been used with some success in separating process components on other reaction time tasks (e.g., Sternberg, 1977), we were able to independently estimate the duration of each of these processing stages.

We also administered two other versions of the task in which the only difference was that subjects were required to decide whether (a) two digits were the same with respect to oddness or evenness, or (b) two letters were the same with respect to vowelness or consonantness. The data analysis addressed two questions regarding generality. First, were parallel measures of stage duration (estimates derived from separate blocks of items) more highly inter-correlated than correlated with other stage durations? This is a direct test of stage independence. Second, were stage durations estimated from tasks with different content (words, digits, or letters) more highly inter-correlated or were alternative stages taken from same-content tasks more highly inter-correlated? This is a direct test of the relative importance of content and process. Although the analyses were rather complex, the general finding was that processes were somewhat independent, and also general across contents. That is, fast encoders were not necessarily fast comparers, but fast encoders on the word task were also fast encoders on the digit task.

One of the problems with this approach to studying dimensionality is that it relies on a model of performance that assumes serial execution of processing stages. In our more recent work (Kyllonen, Tirre, & Christal, 1988), we have relaxed this assumption by applying both those models that assume serial execution and those that do not in estimating stage durations. (We also have abandoned the precueing technique because its validity depends on the serial execution assumption.) Following Donaldson's (1983) analysis, stage durations can be estimated in two ways. Assume an ordered set of tasks, each of which can be characterized as requiring a proper superset of the processes of its predecessor. For example, the following set of tasks, each of which requires processing a pair of words, might be characterized this way. reaction time, choice reaction time, physical matching, name matching, semantic (meaning) matching. That is, reaction time consists only of a *reaction* component; the choice task adds a *decision* component, the physical matching task adds *comparison*, name matching adds *retrieval* from long-term-memory, and semantic matching adds *search* through long-term memory.

One can estimate each of these stage durations either by subtracting latency on the predecessor task from latency on the target task (the difference score model), or by statistically holding constant the duration of all predecessor tasks (the part correlation model). The two models employ differing assumptions about the relationships among task components. The difference score model assumes nothing about the relationship between the duration of the target component (e.g., comparison) and the duration of the predecessor task (e.g., choice reaction time). Thus, this correlation is a parameter to be estimated. But the cost of this flexibility is the assumption that the duration of the target component (e.g., comparison) remains constant, regardless of whether the component is embedded in the physical matching task, the name matching task, or whatever. Conceptually there are two problems with this assumption. Consider the reaction component. It may be that reaction is rapid when nothing else is going on, as on the simple reaction time task, but slow when it follows complex processing, as on the semantic matching task. Or it could be the opposite, due to parallel processing: Reaction appears slow on the simple reaction time task because it is the only process executing; but on the meaning identity task, the reaction begins before decision ends, and thus appears fast (as is specified in process cascading models, McClelland, 1979).

The part correlation model avoids this assumption and allows for variability in stage durations over different tasks. This is represented as freedom in the regression weight associated with stage duration to differ from 1.0. But in order to achieve this flexibility, the part correlation model must compensate with an assumption not required with the difference score model. In the part correlation model, it is assumed that the duration of the target stage is uncorrelated with the duration of the predecessor task. For example, the duration of the comparison component in the context of the physical matching task would be assumed to be uncorrelated with response time on the choice reaction time task.

Which of these sets of assumptions is correct, those associated with the part correlation model or those associated with the difference score model? It is not possible to tell, but it is possible to employ both models and then to be confident of relationships only when the models agree.

We took this approach in attempting to estimate the relationship between processing stage durations and performance on a vocabulary test, and also on a paired-associates learning task. Vocabulary is an interesting test case because it is a good measure of general intelligence. The current view is that breadth of word knowledge reflects efficient learning processes in inferring word meanings in context (Marshalek, 1981; Sternberg & Powell, 1983). An additional motivation for looking at vocabulary as a criterion was that a considerable literature has evolved from Hunt and colleagues' (Hunt et al., 1973) early finding of a relationship between the duration of the retrieval stage (as estimated by the difference between response time on the name and physical matching tasks) and verbal ability.

Contrary to Hunt et al. and other previous work, however, we did not find much of a relationship between *retrieval* speed and vocabulary ($r = .17, N = 710$), but we did find a strong relationship between *search* speed and vocabulary ($r = .49$). Subjects capable of quickly accessing semantic attributes of words, controlling for how quickly they did other kinds of information processing, had larger vocabularies than did other subjects.

We found a similar relationship between processing speed and learning, but only in particular circumstances--namely, when study time on the learning task was extremely short (.5 to 2 seconds per pair). The component analysis again made it possible to isolate the semantic search component, as opposed to other processing speed components, as the one consistently most critical in determining learning success. Over a number of studies (which varied on block size, recognition vs. recall responses, etc.), the correlation between learning success and response time on the meaning identity test, controlling for (or eliminating by subtraction) response time on other information processing tests, ranged from $r = .30$ to $r = .50$. In some studies, other information processing speed components predicted learning outcomes, but only inconsistently.

We currently are engaged in two lines of extension to the processing speed work. One is motivated by the idea that information processing speed may be closely tied to working memory capacity insofar as both measures reflect the dynamic activation level of a memory trace (Woltz, 1987). An intriguing implication of this idea has to do with individual differences in the maintenance of activation. In most learning tasks, we do not simply access a term once and only once. Rather, there is redundancy in instructional materials, which allows for multiple accesses of a concept in an instructional episode. Thus, the important search speed variable is not merely how quickly a concept can be accessed on first encounter, but also how quickly the concept can be re-accessed on second, third, and fourth encounters. Woltz (1987) has shown not only that subsequent accesses are much faster than first encounters, but that there are substantial individual differences in the amount of improvement in speed from first to subsequent encounters. Interestingly, those who benefit most are not necessarily those who are quickest initially. We explore further ramifications of the idea of activation as a concept underlying working memory capacity in the next section.

A second extension to the processing speed work involves the exploration of reaction time distributions as a way of determining how subjects process items. There is some work (Hockley, 1984; Ratcliff & Murdock, 1976) suggesting that reaction time on simple tasks actually reflects two underlying components: a normally distributed processing component (e.g., true comparison time) and an exponentially distributed waiting time component (e.g., time of attention lapses and the like). We are currently investigating the feasibility of estimating these reaction time components and determining whether they reflect reliably different processes (Fairbank, in preparation).

In summary, we are continuing to explore a number of mathematical models for identifying component processing speed, and for determining the relationships among different kinds of processing. One benefit from this kind of analysis is that it enables the determination of whether processing speed is a single construct or whether there are multiple varieties of processing speed (the latter appears to be the case). The implication for test development has to do with how, and how many different kinds of tests will be necessary, to measure processing speed.

A second benefit from this kind of analysis is that it allows one to determine what kind of cognitive processing affects learning (in different contexts). One result is that it appears that general reaction speed is not as highly related and therefore fundamental to learning as might be expected on the basis of work by Jensen and others. We have found relationships between basic reaction time and learning, but the particular component of speed of searching semantic memory appears to be the more critical predictor of verbal learning success. This is shown both in studies employing vocabulary scores as a criterion and in those employing a highly speeded presentation of material to be learned. (Perhaps both tasks reflect the learner's ability to quickly elaborate on the stimulus material.)

Processing Capacity

Although much of the early work on the project was concerned with response time, we recently have begun focusing more attention on similar kinds of analyses of working memory capacity. It now appears, not only on the basis of our own work (Kyllonen, Stephens, & Woltz, 1988; Woltz & Christal, 1985) but on work from a number of laboratories (Anderson & Jeffries, 1985; Daneman & Carpenter, 1980; Hitch, 1978), that this component of the information processing system is responsible for learner differences on a wide variety of learning tasks.

In keeping with contemporary views of the human cognitive architecture, we propose that working memory may be defined as that portion of memory currently in a highly active or accessible state; that is, whatever is being processed or attended to at any given time. The individual differences corollary is that greater working memory capacity should be associated with greater attentional and learning capabilities. Woltz (1987) has pointed out that this quite general description of working memory capacity is realized in the literature in two rather different forms, which we will refer to as the *processing workspace* and *activation capacity* models.

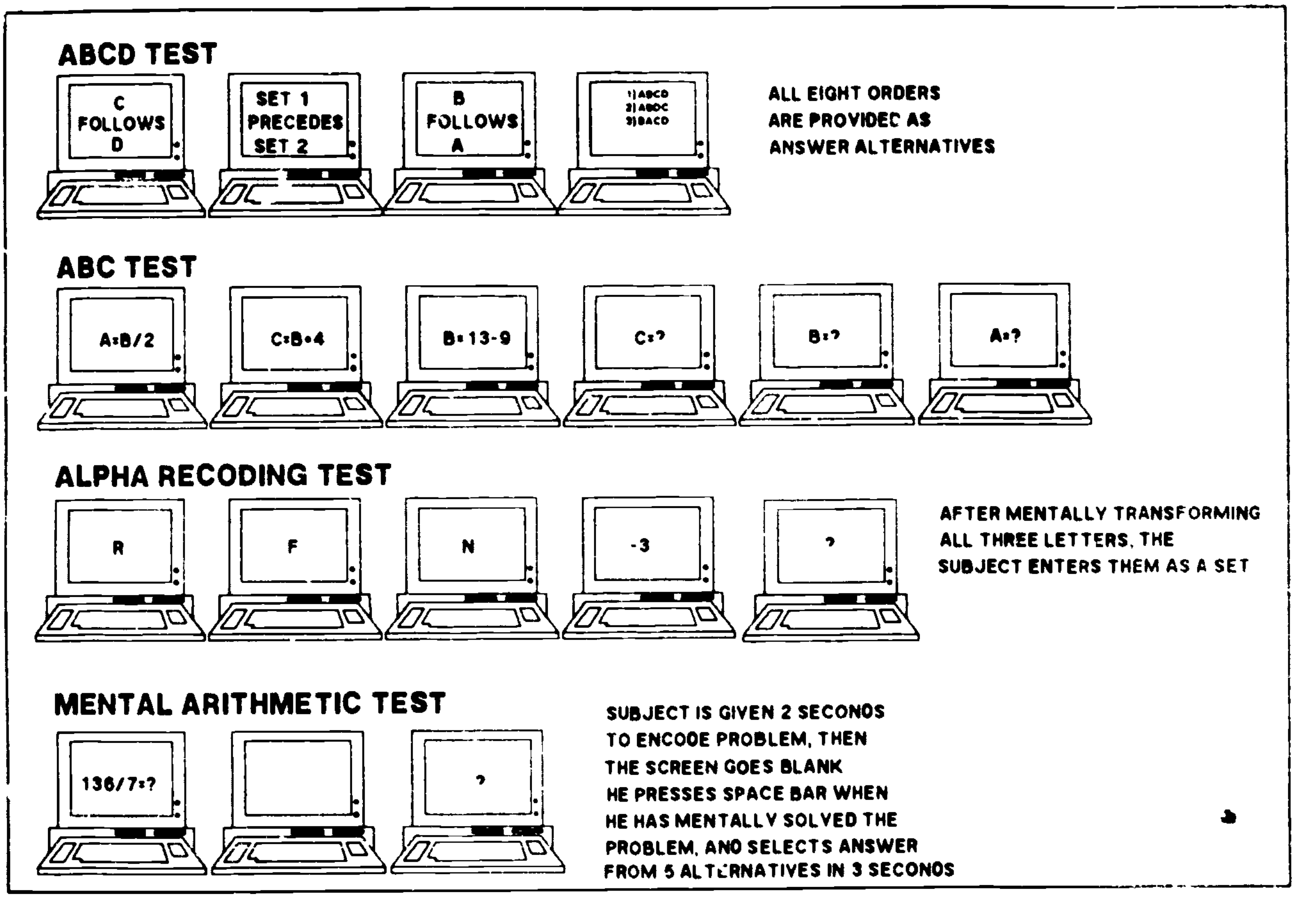
The *processing workspace* model of working memory, due largely to the work of Baddeley and Hitch (1974), proposes a limited, consciously controlled, short-term memory capable of storing roughly three to nine items simultaneously. The capacity of this structure is determined mostly by how efficiently one processes new incoming information. Much of our work on working memory to date has consisted of the application of the processing workspace model to the development of working memory capacity tasks. The guiding construction principle is that the task requires the retention of some information, while simultaneously requiring the processing or transforming of other information. This principle is consistent with Baddeley and Hitch's (1974) original definition, and seems on the surface to lend itself readily to ecologically valid tests of memory capacity insofar as much of learning demands simultaneous retention and processing. In contrast, what is required on span tests seems contrived and not typical of what people actually do when engaged in realistic learning.

Figure 2 shows sample items from various tests developed in our laboratory. In the "ABCD Test," the subject is informed that all items involve two sets of letters. The first set is defined as the letters A and B, and the second set as the letters C and D. The subject is then presented three statement frames that constrain the ordering of the four letters. In the item pictured, for example, the subject is presented a frame which states that C follows D. The subject next is presented a frame which states that Set 1 precedes Set 2. The subject is expected at this point to note that the letters A and B will precede the letters C and D in the final list. On the third frame, the subject is informed that B follows A. The frames are presented successively, and the subject cannot look back to retrieve previous statements. From the three assertions, the subject would be expected to generate the proper ordering of the four letters, ABDC. The test probe is then presented, and the subject responds by selecting one of the eight orders presented as multiple-choice alternatives.

A second test, the "ABC Test," also involves successive presentation of instruction frames; only here, the instruction frames are assignments of either values (e.g., A = 3), expressions (e.g., A = 24 - 17), or equations (e.g., A = B / C). In the item pictured, the subject first sees that A gets the value of B divided by 2. The subject does not yet know what B is and so must remember the equation. The next frame states that C gets the value of B plus 4. Again, the subject still does not know the value of B and so must remember the equation. Finally, the subject is shown that B is 13 minus 9, and this allows him or her to solve for C and A. But in order to do so the subject must remember the equations for C and A. The subject is then tested for which values he or she can remember.

In the third test, the "Alpha Recoding Test," the subject is shown either one, two, or three random letters, one at a time on successive frames. On the next frame the subject is instructed either to add or

EXAMPLE ITEMS FROM TESTS MEASURING ATTENTION CAPACITY


ABCD TEST

ALL EIGHT ORDERS ARE PROVIDED AS ANSWER ALTERNATIVES

ABC TEST

ALPHA RECODING TEST

AFTER MENTALLY TRANSFORMING ALL THREE LETTERS, THE SUBJECT ENTERS THEM AS A SET

MENTAL ARITHMETIC TEST

SUBJECT IS GIVEN 2 SECONDS TO ENCODE PROBLEM, THEN THE SCREEN GOES BLANK HE PRESSES SPACE BAR WHEN HE HAS MENTALLY SOLVED THE PROBLEM, AND SELECTS ANSWER FROM 5 ALTERNATIVES IN 3 SECONDS

Figure 2. Sample Test Items Measuring Working Memory Capacity. Test results were analyzed in Christal (1987).

to subtract 1, 2, or 3 (*n*). Add and subtract in this context means to determine which letter follows or precedes each of the target letters by *n* positions. After mentally recoding all the letters, the subject presses the space bar and enters the answer. The other test shown in Figure 2, the "Mental Arithmetic Test," is self-explanatory.

As with information processing speed, an important initial question to be asked regarding performance on these kinds of tasks is whether working memory capacity is a unitary or multidimensional construct. A related question concerns the relationship between working memory and performance on other more conventional aptitude tests. We addressed both questions in a large-scale correlational study recently completed (Christal, 1987). We administered the tests shown in Figure 2, along with additional measures such as Memory Span, the AB Sentence-Picture Verification Test[2] (Baddeley, 1968), and the Sunday-Tuesday Test[3] (Hunt et al., 1973). Additionally we had available subjects' scores on the Armed Services Vocational Aptitude Battery (ASVAB), which consists of 10 paper-and-pencil subtests, such as Word Knowledge, Paragraph Comprehension, Numerical Operations (Number Facts), and General Science Information.

A correlation matrix was generated from the percent correct and the latency scores on the computerized tests and the raw scores on the timed ASVAB subtests. A principal-axis factor analysis of this matrix yielded four factors. A Working Memory factor was defined primarily by percent correct scores from the ABC Test ($r = .80$), but also was heavily loaded by the ABCD Test, Mental Arithmetic Test, and the other working memory measures (all of which showed $r > .60$). The two verbal measures, Word Knowledge and Paragraph Comprehension, had only modest loadings on this factor (r $< .15$). In addition to the Working Memory factor, separate Verbal and Speeded-Quantitative factors were extracted. The Verbal factor was defined by ASVAB Word Knowledge ($r = .77$), but also was highly loaded by both the ABCD Test and the AB Sentence-Picture Verification Test, which may be thought of as an abridged version of the ABCD Test ($r > .50$). The Speeded-Quantitative factor was

---

[2] This test requires subjects to judge whether a sentence such as "A is not preceded by B" matches a string such as "BA."

[3] This test requires subjects to perform base 7 addition on days-of-the-week values, with Sunday assigned 1, e.g., "Sunday + Tuesday = Wednesday."

defined by the Numerical Operations subtest ($r = .75$), but it also was significantly loaded by latencies from the Mental Arithmetic Test and the Sunday-Tuesday Test ($r > .30$). The basic pattern of results found here has been corroborated in a recently completed follow-up study.

Taken together, the results suggest the involvement of both domain knowledge (quantitative and verbal) and a domain-independent working memory in memory test performance. In addition, it appears from the data over the two studies that the Working Memory factor subsumes the Reasoning factor. That is, individual differences in reasoning proficiency may be due entirely to differences in working memory capacity. Christal notes that the factor on which all the reasoning tests in the battery loaded highly is a Working Memory factor in that the test that defined it, Alpha Recoding ($r = .68$, in the follow-up study), does not appear to involve reasoning per se but clearly depends on working memory capacity.

Recently, we have begun investigating an alternative to the processing workspace model which is based on a different conceptualization of working memory. The activation capacity model, based primarily on Anderson's (1983) ACT* theory, defines working memory, not as a separate short-term store but rather, as a state of fluctuating activation patterns characterizing traces in long-term memory. According to this theory, long-term memory is a network of traces, each characterized by resting activation levels. Traces become activated when they become the focus of attention, or are linked to the focus of attention, then fade into a state of deactivation as other traces move to the center of focus. Working memory is said to be a "matter of degree" rather than an all-or-none state, in that at any given moment, a trace might be the focus of attention (and thereby be at a peak activation level) or it might be continuously fading from attention if, for example, it was the focus a few seconds earlier.

The application of this model has resulted in tests of working memory capacity that look quite distinct from those based on the processing workspace model. Figure 3 illustrates a test developed by Woltz (1987) to reflect individual differences in activation capacity. In this test, subjects are presented a series of word pairs and are requested to determine whether or not the words are synonyms. Occasionally, words are repeated one, two, four, or eight items later. As Figure 3 shows, mean

EXAMPLE ITEMS

fate	destiny
humid	damp
complain	thunder
humid	damp
polite	courteous
polite	kindle
astonish	unstable
conquer	arrange
visitor	guest
vacant	empty
complain	gripe

MEASURES OBTAINED

1. Verbal Information Processing Speed

M= 1295 ms; SD=326 ms

2. Residual Activation Strength

Lag of Repeated Item	Mean Savings	S.D. Savings
1	181 ms	218 ms
2	124 ms	229 ms
3	108 ms	214 ms
4	107 ms	219 ms

Figure 3. Woltz's (1987) Procedure and Resulting Statistics for Measuring Memory Activation Capacity.

response time is 1265 ms if neither of the words was shown before, but that time is reduced by 191 ms if one of the words was encountered on the previous item, and by 107 ms if one of the words was encountered eight items ago. The interpretation is that the word encountered even eight items ago is still more highly active than it would be at its true resting state, and therefore is processed faster. Woltz argues that individual differences in the response time facilitation effect reflect differences in activation capacity.

Given that we can define working memory capacity in two distinct ways, an important next question is: What is the empirical relationship between the two kinds of measures, and even more importantly, what is their relationship to learning? Cognitive analyses of learning tasks (Anderson, 1987; Anderson & Jeffries, 1985), such as mathematics learning or learning a computer programming language, suggest that the limiting factor in learning is the working memory bottleneck. But the proof of this assertion is often rather theoretical, based on a rational analysis of learning task requirements, supplemented by a formal computer simulation of learning processes. An individual differences analysis of the role of working memory in learning can be a useful supplement to this kind of formal analysis, and is a fair test of the theoretical claim (Underwood, 1975). Thus, we have recently begun investigating the relationship between working memory capacity (as measured by tests such as those displayed in Figures 2 and 3) and performance in realistic learning contexts. We currently are investigating the acquisition of electronics troubleshooting (Kyllonen, Stephens, & Woltz, 1988) and computer programming skills (Kyllonen, Soule, & Stephens, 1988) and other procedural learning tasks (Woltz, 1987). In all cases, we find that working memory, as indicated by both the processing workspace and activation capacity measures, is a strong predictor of learning outcomes. These analyses are beginning to clarify our understanding of working memory. These studies also suggest that the particular tests of working memory capacity that we have already developed (Figures 2 and 3) are solid candidates for inclusion in future testing batteries.

Knowledge

In our four-source framework for cognitive skill assessment, we refer to declarative knowledge and procedural skills as *enablers*. It has been argued that the main contribution from cognitive psychology to the new generation of psychological tests is in how we now can assess the *mediators*--information processing speed and working memory capacity--rather than the enablers. The idea behind this thinking is that existing tests already do an adequate job at sampling the breadth of an individual's knowledge. For example, existing vocabulary tests probably are fair samples of what a person knows (although faceted vocabulary tests with a consistent sampling scheme are probably even better, Anderson & Freebody, 1979; Cronbach, 1942; Marshalek, 1981). Also, the ASVAB includes a number of subtests--Auto and Shop Knowledge, Mechanical Comprehension, Electrical Knowledge--that are clearly designed to sample the breadth of technical knowledge a student brings to the test.

Thus, in much of our research, the measurement of knowledge has played a rather small role, especially when considered against the backdrop of its critical role in current cognitive theories generally. In experiments conducted to date, we have assessed knowledge primarily as a means for statistically controlling its effects; our main goal has been to investigate the mediator variables, which is best done by holding the knowledge effect constant.

Perhaps the reason we have failed to progress in assessing the role of knowledge in learning is that our learning tasks have purposely been rather domain-independent. It may be that advances in understanding the role of knowledge will be forthcoming only once we begin our actual complex learning experiments (described in the next section). Still, there has been a considerable body of cognitive research conducted over the last 10 years that enables speculations.

We propose that an individual's declarative knowledge base may be characterized along four general dimensions: *depth*, *breadth*, *accessibility* (durability), and *organization*. *Depth* refers to the amount of domain-specific conceptual knowledge possessed by the individual. Conventional achievement tests, and especially job surveys as they are employed in assessing trainee or apprentice status, are designed to tap this dimension of declarative knowledge. *Breadth* refers to the amount of general factual knowledge available. Current intelligence tests, such as the Wechsler Adult Intelligence Scale (WAIS), include an Information subtest designed to probe breadth of knowledge. Vocabulary tests can also be seen as measures of breadth of knowledge. *Accessibility* refers to the strength of the knowledge; that is, the likelihood (and the speed with which) it will be accessed in a situation in which it could be used. Accessibility is both a general characterization of all knowledge an individual possesses and a specific parameter of every fact in the knowledge base. Accessibility is also a dynamic property of specific knowledge, in that it weakens with disuse and grows stronger with practice. *Organization* refers to the relations and connections among the facts in the knowledge base. A considerable body of research in cognitive science has grown around the idea that acquiring expertise in a domain involves the reorganization of facts in the domain (e.g., Lesgold, 1984).

Various methods have been developed to tap these knowledge dimensions. Clustering and scaling methods have been used to map the organization of knowledge in numerous domains such as physics (Chi, Glaser, & Rees, 1982), biology (Stephens, 1987), computer science (Adelson, 1981), psychology (Fabricious, Schwanenflugel, Kyllonen, Barclay, & Denton, 1987), and so on. Typically, a student is asked to judge the similarity of two concepts selected from the domain. Clustering and scaling methods are used to capture the underlying model used by the student to generate the similarity judgments.

There are many ways to tap accessibility of knowledge. We have used the sentence verification technique extensively (e.g., Tirre, Royer, Greene, & Sinatra, 1987). Learning in the typical training situation involves listening to a lecture or reading a text, then solving problems based on the material heard or read. The sentence verification technique is designed to probe the amount of material the learner was able to successfully encode and store in long-term memory following the listening or reading episode. The technique requires learners to discriminate between accurate paraphrases of sentences previously read and paraphrases that are inconsistent with what was read. Other techniques such as the cloze procedure (fill-in-the-blanks of sentences extracted from the preceding text) have been used for a similar purpose (Landauer, 1986). We are currently using the sentence verification technique for tracking the accumulation of declarative knowledge during the course of short (45 minutes) instructional episodes in computer programming (Kyllonen, Soule, & Stephens, 1988) and electronics troubleshooting (Kyllonen, Stephens, & Woltz, 1988).

Even the measurement of the depth and breadth dimensions of knowledge may benefit from recent work in cognitive science. The most innovative recent developments in probing declarative knowledge have been pursued by researchers concerned with achievement testing (Frederiksen, Lesgold, Glaser, & Shafto, in press; Glaser, Lesgold, & Lajoie, in press; Haertel, 1985; Lesgold et al., 1987). Glaser et al. point out that current methods, typically 5-alternative multiple-choice tests, suffer two key drawbacks. First, the alternatives cannot possibly accommodate all the possible misconceptions a student could possess, and thus are of limited diagnostic utility. Second, the alternatives may give away the answer, as has been shown in other realms.

Glaser et al. discuss the potential of cognitive approaches to knowledge assessment, which in contrast rely primarily on a very detailed analysis of verbal protocols extracted from students struggling with new material or applying what they have already learned. Analysis of these kinds of protocols has played a critical role in the development of a cognitive science (Ericsson & Simon, 1984) and serves as the primary basis for what Glaser, Lesgold, Lajoie, et al. (1985) have dubbed *cognitive task analysis*. The problem with wholesale adoption of the technique at this time is expense. Protocol analyses are costly in both subject and interviewer time, and are therefore not appropriate for inclusion in a test battery.

But Glaser et al. suggest an ingenious compromise between conventional and protocol methods. In their *hierarchical menus methodology*, students select alternatives from a series of linked menus. For example, if there are five alternatives to each menu and there are three levels of linked menus, there can be $5^3 = 125$ response alternatives. This is superior to simply presenting 125 alternatives on screen, for two reasons. First, selecting from among 125 alternatives would impose a severe processing load on subjects, and would inv[illegible] nuisance individual-difference variation in strategy selection and test-taking strategy. Second, the hierarchical arrangement can closely mirror the way in which a student is thinking about a problem, in a kind of top-down fashion.

Thus far, this approach to probing an individual's knowledge has been employed in one of the CLASS tutoring systems. Bridge (Bonar & Cunningham, 1986), which teaches learners how to program in Pascal, presents general programming problems to be solved. At the top level (the first set of questions), the alternatives are general categories or general approaches to the problem (e. g., "add something together" or "keep doing something"). Once the student selects a category, he or she is presented a list of alternatives that refine the category selection, and so on, until a fully specified answer is selected From pilot testing using Air Force subjects, the method has proved general enough to accommodate the vast majority of potential responses to particular programming problems; therefore, the approach seems highly promising as a way of assessing knowledge status in the student.

To summarize, although we have not yet fully explored the domain of how to probe a learner's declarative knowledge base, we have made some important initial steps. It is likely that as we begin further testing in the more complex tutoring systems environments, the methods described in this section will be refined further.

Skills

We define skills or *procedural knowledge* as it is referred to in the cognitive science literature, fairly informally, as any unit of knowledge that is typically or would likely be represented in production system simulations in the form of an if-then rule or series of if-then rules. This is any knowledge or skill the student has that might bear directly on problem solving ("how-to knowledge"). Procedural skill varies widely along the generality dimension; at the most general level are problem-solving heuristics or approaches, such as working backward, means-ends analysis, or persisting in the face of uncertainty. At the opposite end of the continuum are very specific procedures, such as moving the cursor to position 12, 45 when required to delete a character at position 12, 45.

One fairly consistent finding in cognitive research is that although specific procedures are trainable, general procedures are quite resistant to modification. This finding is certainly not due to a shortage of attempts to modify general skills. Kulik, Bangert-Downs, and Kulik (1984) reviewed over 50 studies of the effects of extensive coaching for the Scholastic Aptitude Test (SAT). They concluded that the effects, even for long-term training, were quite small (approximately one-sixth to one-third standard deviation, or 17 to 34 points). The results of Venezuela's Project Intelligence (Herrnstein, Nickerson, de Sanchez, & Swets, 1986) may be seen similarly as somewhat disappointing. Despite an ambitious project in which domain-free thinking skills were taught 4 days per week, in 45-minute lessons, for an entire year, the actual changes experienced on standard measures of cognitive skill (intelligence tests) were quite minuscule (about .3 sd). These findings should not have come as any great surprise. Attempts to have students transfer general problem-solving approaches to superficially distinct but isomorphically identical problems have repeatedly failed (e.g., Brown & Campione, 1978; Simon & Hayes, 1976).

On the other hand, there is good evidence for the modifiability of specific skills, especially in context. Schoenfeld (1979) has shown how training in mathematical heuristics (e.g., draw a diagram, simplify the problem, test the limiting case) can facilitate subsequent problem solving so long as the instruction is wedded tightly to the domain material simultaneously being taught. Recent analyses of transfer of training have shown that skill transfer is excellent and quite predictable when the skills transferred are related at some conceptual level to the new skills (Anderson, 1987; Kieras & Bovair, 1986).

The implications of these two results for testing purposes are apparent. On the one hand, specific procedural knowledge is rather easily modifiable and therefore ought to perhaps be trained rather than tested for, at least in the personnel selection and classification context. Recent work on diagnostic monitoring (Frederiksen et al., in press; Lesgold et al., 1987) shows how tests can be used to tailor instruction and are thus appropriate for this purpose. On the other hand, general procedural knowledge should have an important predictive relationship to learning ability, and it seems to be fairly immutable. General procedural knowledge, therefore, is an ideal capability to test for in entrance (selection and classification) testing. It is interesting that researchers from very diverse perspectives--psychometric (Cattell, 1971), information processing (Sternberg, 1981a), and artificial intelligence (Schank, 1980)--have argued consistently for the importance of the ability to cope with novel problems as a key aspect of intelligence, and therefore as an ideal candidate for inclusion in aptitude test batteries.

Do we now test for general procedural knowledge, or general problem-solving skills? As was the case with declarative knowledge, there certainly are in existence paper-and-pencil tests that would appear to tap very general problem-solving skill--Raven's Progressive Matrices being an excellent example. And about 7 years ago, ETS began supplementing its existing Verbal and Quantitative portions of the Graduate Record Examination with a new test of Analytic ability (Wilson, 1976). The ASVAB comes close to testing general problem-solving ability with the Arithmetic Reasoning subtest. This subtest consists of story problems such as "How many 36-passenger buses will it take to carry 144 people?" (DoD, 1984). Recall that the Arithmetic Reasoning subtest loaded highly on the Working Memory factor in the Christal (1987) study, which suggests an intriguing research question: What is the relationship between working memory and procedural skill?

We can think of working memory capacity as mediating the development and efficiency of general problem-solving strategies. But an alternative view of the relationship between the two constructs assigns the central role to working memory. Baddeley (1987) has proposed a model of working memory consisting of various slave storage subsystems (for storing linguistic information, spatial information, etc.), along with a central executive which monitors and coordinates the activities of the subsidiary storage systems. Executive skill, then, is skill in monitoring one's problem-solving processes, adapting to changing task requirements, successfully executing general problem-solving strategies, allocating resources where they are needed, and more generally, changing processing strategy in accordance with changes in processing demands.

In this way, the executive can be seen as the most important component of working memory. Yet, though we have a reasonable understanding of how the subsidiary storage systems function, according to Baddeley the workings of the central executive still remain largely a mystery. An important and exciting research direction is to begin devising means for measuring executive skill and thereby begin unraveling that mystery.

Modeling Learning Skills

Learning Skills Taxonomy

If we can adequately measure knowledge and the various skills associated with the four sources, an important next step in the research program is to demonstrate the relationship between those scores and scores generated from a trainee's interaction with a learning task. We believe that learning should be expressible in terms of (i. e., predictable from) the underlying components, but it is necessary to prove that this is the case.

Much of our research until fairly recently has used grossly simplified learning tasks as criterion measures against which to validate the new cognitive abilities measures. For example, in the Kyllonen-Tirre-Christal (1988) study, performance on various paired-associates tests were used as criteria; and in other studies, we have employed comparably simple, short-term learning tasks. The logic underlying this decision is twofold. First, we are concerned with developing rigorous models of the aptitude-learning-outcome relationship; and simple, short-term learning tasks afford more control over the instructional environment. But second, we believe that the kind of learning involved in even these simple tasks is at some fundamental level the same as that involved in more realistic learning situations. Or, conversely, even apparently complex classroom learning can be analyzed and decomposed into a series of much simpler learning acts.

If we accept the notion that even complex learning tasks can be broken down into their constituent learning activities, then it obviously would be useful to specify the nature of those basic learning activities. One proposal that has been useful in our work, based largely on Anderson's (1987) three-stage model of skill acquisition, is represented on the right side of Figure 1. The idea is that cognitive skills develop through an initial engagement of declarative learning processes ("memorizing the steps"), followed by an engagement of proceduralization processes ("executing the steps"), then finally refinement processes ("automatizing the steps"). As Figure 4 shows, different performance measures will be sensitive to the course of skill development at various points along the way. When first learning a skill, many mistakes will be made, and accuracy measures will be the most sensitive indicators of skill development. Later, when the skill is known, few mistakes will be made, and performance time measures will be the most sensitive indicators. Still later, performance time will approach a minimum as the target skill becomes increasingly automatized, but there might still be considerable variability in whether (and how much) other processing can be occurring while the target skill is being executed.

We (Kyllonen & Shute, in press) recently elaborated on this simple taxonomy in proposing that in addition to the status of the skill (i.e., whether the skill is in a declarative, procedural, or automatic state, which we identified as the *knowledge-type* dimension), learning could be classified along three other dimensions: the *learning environment*, the *domain*, and the learner's *cognitive style*.

The *learning environment* specifies the nature of the inference process required by the student: The simplest learning act involves rote memorization. Learning by actively encoding, by deduction, by analogically reasoning, by refinement through reflection following practice, by induction from examples, and by observation and discovery involves successively more complex processing on the part of the learner. The second dimension, the resulting *knowledge-type*, as indicated above, specifies whether the product of the learning act is a new chunk of declarative knowledge (a new fact or body of facts) or new procedural knowledge (a rule, a skill, or a mental model). The third dimension, the *domain*, refers to whether learning is occurring in a technical, quantitative domain or a more verbal, non-technical domain. Together, these three dimensions specify a particular kind of learning act. The fourth dimension, the learner's *cognitive style*, is a property of the learner rather than of the instructional situation per se. But we included it in recognition of the possibility that we cannot be

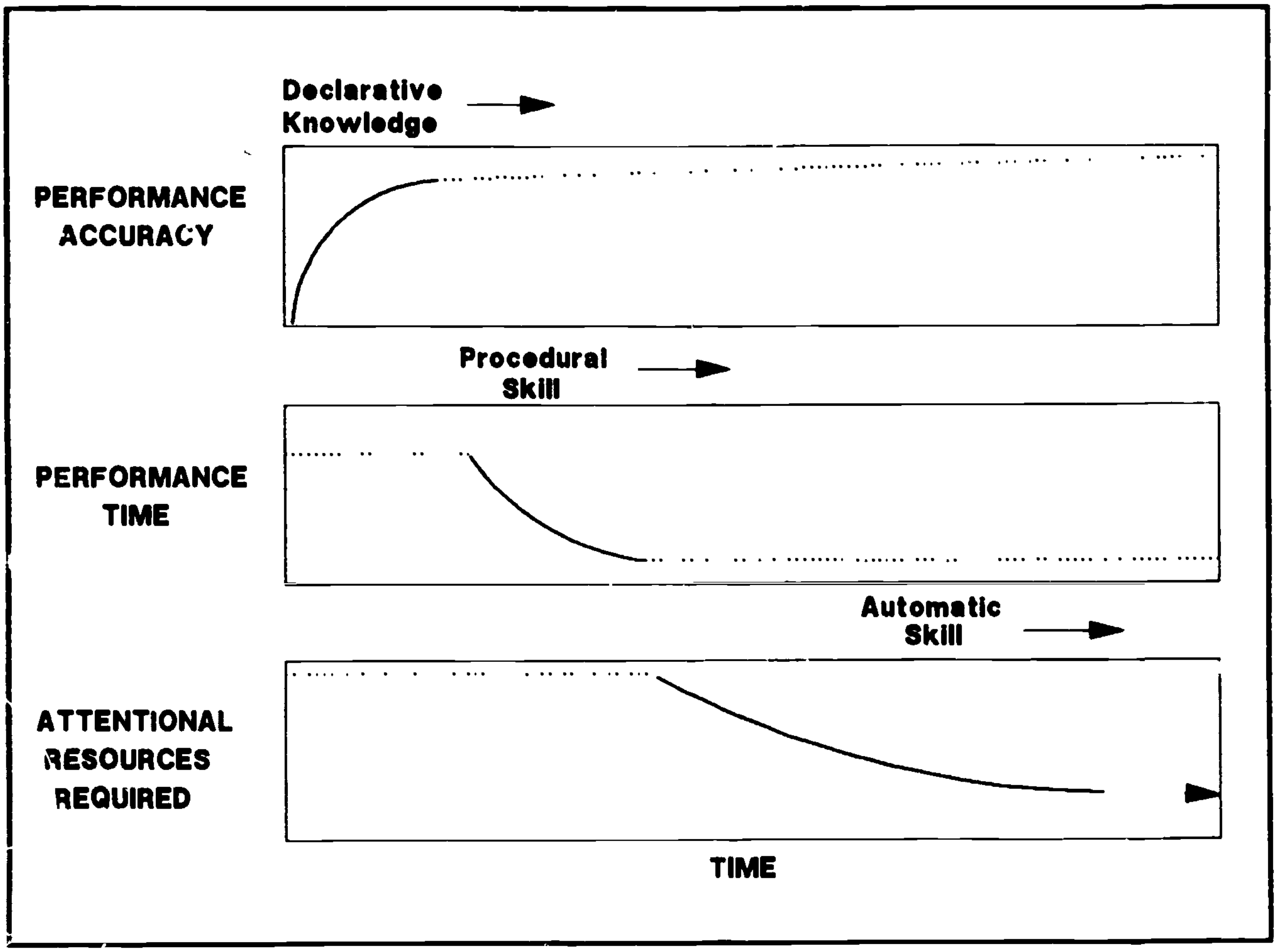


Figure 4. Performance Curves for Three Dependent Measures as a Function of the Stage of the Skill Being Measured. The different dependent measures are optimally sensitive to individual differences at different stages.

certain on any task of what learning skill is being assessed unless we consider how the learner is approaching the task.

Our proposal, which has not in any sense been put to the test, is that the taxonomy should prove useful in two ways. First, it provides a sampling space from which we may draw learning tasks. The goal of the LAMP effort is to model learning ability using cognitive skill measures; the taxonomy specifies the range of learning tasks for which we must develop adequate models. Second, in reverse fashion, the taxonomy specifies the kinds of micro-level learning acts that combine to make complex learning. This aspect provides a task analysis tool. Our idea is that we can inspect the requirements of any complex learning situation, in the classroom or in front of a computer, and specify what learning acts are occurring. Given any instructional exchange, we can find a cell in the taxonomy that represents that exchange.

Complex Learning Assessment (CLASS)

One potential stumbling block for any program like ours is that it is not easy to monitor progress. To determine whether our innovative measurement methods are valid predictors of learning success, it is necessary to observe students engaged in learning. Two approaches have traditionally been taken. One is to validate the new tests against some criterion reflecting success in operational training, such as final course grade point average. The benefit of this approach is that inferences from the research are direct, but there are a number of drawbacks: Data collection is extremely slow, instructor quality is highly variable and may interact with learner characteristics in affecting learning outcomes, and there is no allowance for manipulating the learning task in any way so as to allow "what-if" questions regarding validity (e.g., "what if the instructor encouraged more questions, would that differentially affect student outcomes?").

The second approach is to simplify the learning task such that it is under the experimenter's control and can be administered within a single session. With complete control over the learning task, one can ask and test what-if questions easily. Unfortunately, in so modifying the learning task, the researcher cannot necessarily continue to assume that the instruments shown to be valid in the experimental context will prove to be valid in predicting success in more realistic learning situations.

Our solution to the validity problem represents a compromise between these two positions. We are currently designing intelligent computerized tutoring systems to teach computer programming, electronics troubleshooting, and flight engineering in 56-hour mini-courses (Learning Research & Development Center, 1987). In addition, we will add new mini-courses over the next several years. The tutoring systems are being designed to produce a rich variety of indices of the learner's curriculum knowledge and his or her progress in acquiring the new knowledge and skills being taught. The tutoring systems are sufficiently flexible so that it is easy to modify the instructional strategy and thus ask what-if questions. The learning involved, however, is not trivial. It has been estimated that 1 hour of tutored instruction is equivalent to approximately 4 hours of regular classroom instruction (Anderson, Boyle, & Reiser, 1984); thus, these mini-courses are quite extensive. A major goal of our current research efforts is to use the taxonomy to generate the most expressive indices of the student's learning experience.

We envision a broad range of research questions that can be addressed once we begin gathering data with these kinds of learning indices. First, the indices can serve as alternatives to end-of-course achievement test scores as criteria for validating new cognitive aptitude tests. An index such as "probability of remembering an instructional proposition (as a function of the amount of study and presentation lag)" is more precise and potentially more general than a broad achievement test score. Such a fine breakdown of the learning experience also permits enhanced analyses among the indices themselves. For example, we can begin investigating more precisely questions concerning the relationship between initial knowledge acquisition and the subsequent ability to turn that knowledge into problem-solving skill, or the ability to tune that skill with more problem-solving experience.

Finally, developing rich profiles of an individual learner's strengths and weaknesses in the form of elaborate assemblies of learning indices should permit a reassessment of the aptitude-treatment-interaction (ATI) idea (Cronbach & Snow, 1977). Probably, the inconclusiveness of past ATI research can be traced to the employment of global aptitude indices and global learning outcome measures along with pragmatic limitations on instructional variation. The tutoring systems being developed overcome these limitations by generating richer traces of a learner's path through a curriculum, and by being sufficiently flexible to allow potentially unlimited variations in how instruction is presented.

IV. SUMMARY AND CONCLUSIONS

This paper has outlined some of the research activities underway as part of the Air Force's Learning Abilities Measurement Program (LAMP). The major goal of the project is to devise new models of the nature and organization of human abilities, with the long-term goal of applying those models to improve current personnel selection and classification systems.

As an approach to this ambitious undertaking, we have divided the activities of the project into two categories. The first category is concerned with identifying fundamental learning abilities by determining how learners differ in their abilities to think, remember, solve problems, and acquire knowledge and skills. From research already completed, we have established a four-source framework that assumes that observed learner differences are due to differences in information processing efficiency; working memory capacity; and the breadth, extent, and accessibility of conceptual knowledge and procedural and strategic skills.

The second category of research activities is concerned with validating new models of learning abilities. To do this, we are building a number of computerized intelligent tutoring systems that serve as mini-courses in technical areas such as computer programming and electronics troubleshooting. A major objective of this part of the program is to develop principles for producing indicators of student learning progress and achievement. These indicators will serve as the learning outcome measures against which newly developed learning abilities tests will be evaluated in future validation studies. The indicators also will be applied in studies that investigate the dynamics of knowledge and skill acquisition and in studies that attempt to optimize instruction so as to capitalize on and compensate for learner strengths and weaknesses.

REFERENCES

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition*, *9*, 422-433.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: MIT Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, *94*, 192-210.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1984). Intelligent tutoring systems. *Science*, *228*, 456-462.

Anderson, J. R., & Jeffries, R. (1985). Novice LISP errors: Undetected losses of information from working memory. *Human-Computer Interaction*, *1*, 107-131.

Anderson, R. C., & Freebody, P. (1979). *Vocabulary knowledge* (Tech. Rep. No. 136). Champaign: University of Illinois, Center for the Study of Reading.

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, *10*, 341-342.

Baddeley, A. D. (1987). *Working memory*. New York: Academic Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Advances in learning and motivation (Vol. 8)*. New York: Academic Press.

Bonar, J. G., & Cunningham, R. (1966). *Bridge: An intelligent tutor for thinking about programming* (Tech. Rep.). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Brown, A. L., & Campione, J. C. (1978). Memory strategies: Training children to study strategically. In H. Pick, H. Lebowitz, J. Singer, A. Steinschneider, & H. Stevenson (Eds.), *Application of basic research in psychology*. New York: Plenum.

Carroll, J B. (April, 1981). *Individual difference relations in psychometric and experimental cognitive tasks* (Tech. Rep. No. 163). Chapel Hill: University of North Carolina.

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

Chi, M. T. H., Glaser, R., & Rees, E., (1982). Expertise [illegible] [illegible]lem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. [illegible], NJ: Erlbaum.

Christal, R. E. (1981). *The need for laboratory research to improve the state-of-the-art in ability testing*. Paper presented at the National Security Industrial Association-DoD Conference on Personnel and Training Factors in Systems Effectiveness, San Diego, CA.

Christal, R. E. (1987). [Correlation matrices and factor analysis results from two working memory test batteries]. Unpublished raw data.

Collins, A. (1986). High pay-off research areas in the cognitive sciences. In T. G. Sticht, F. R. Chang, & S. Wood (Eds.), *Advances in Reading/Language Research: Cognitive Science and Human Resources Management*. Greenwich, CN: JAI Press.

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research*, *36*, 206-217.

Cronbach L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

Department of Defense (1984). *Test manual for the Armed Services Vocational Aptitude Battery* (DoD, 1340.12AA). North Chicago, IL: U. S. Military Entrance Processing Command.

Detterman, D. K. (1980). Does "g" exist? *Intelligence*, *6*, 99-108.

Detterman, D. K. (1986, November). *Basic cognitive processes predict IQ*. Paper presented at the Twenty-seventh Annual Meeting of the Psychonomic Society, New Orleans, LA.

Donaldson, G. (1983). Confirmatory factor analysis models of information processing stages: An alternative to difference scores. *Psychological Bulletin, 94*, 143-151.

Embretson, S. (in press). Diagnostic testing by measuring learning processes: Psychometric considerations for dynamic testing. In N. Frederiksen, A. Lesgold, R. Glaser, & M. Shafto, (Eds.), *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1984) *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fabricious, W. F., Schwanenflugel, P. J., Kyllonen, P. C., Barclay, C., & Denton, M. (1987, April). *Developing concepts of the mind: Children's and Adults' Representations of mental activity*. Paper presented at the meeting of the Society for Research in Child Development, Baltimore, MD.

Fairbank, B. B., Jr. (in preparation). *Mathematical analyses of reaction time distributions*. San Antonio, TX: Performance Metrics, Inc.

Frederiksen, N., Lesgold, A., Glaser, R., & Shafto, M. (Eds.). (in press). *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ: Erlbaum.

Glaser, R. (1985, October). *The integration of instruction and testing*. Paper presented at the ETS Invitational Conference, Princeton, NJ.

Glaser, R., Lesgold, A., & Lajoie, S. (in press). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. Glover, J. C. Conoley, & J. C. Witt (Eds.) *The influence of cognitive psychology on testing, Buros/Nebraska Symposium on Testing (Vol. 3)*. Hillsdale, NJ: Erlbaum.

Glaser, R., Lesgold, A. M., Lajoie, S., Eastman, R., Greenberg, L., Logan, D., Magone, M, Weiner, A., Wolf, R., & Yengo, L. (1985, October). *Cognitive task analysis to enhance technical skills training and assessment* (Tech. Rep.). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Haertel, E. (1985). Construct validity and criterion-referenced testing. *Review of Educational Research*, *55*, 23-46.

Herrnstein, R. J., Nickerson, R. S., de Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist*, *41*, 1279-1289.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*, 302-323.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 598-615.

Hunt, E. (1982). Toward new ways of assessing intelligence. *Intelligence*, *6*, 231-240.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (Vol 7)*. New York: Academic Press.

Hunt, E. B. & Pellegrino, J. W. (1984). *Using interactive computing to expand intelligence testing: A critique and prospectus* (Tech. Rep. No. 84-2). Seattle: University of Washington, Department of Psychology.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production system analysis of transfer of training. *Journal of Memory and Language*, *25*, 507-524.

Kulik, J. A., Bangert-Downs, R. L., & Kulik, C-L C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, *95*, 179-188.

Kyllonen, P. C. (1985). *Dimensions of information processing speed* (AFHRL-TP-84-56, AD-A154-778). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Kyllonen, P. C. (1986). Theory-based cognitive assessment. In J. Zeidner (Ed.), *Human productivity enhancement: Organizations, personnel, and decision making* (Vol. 1, pp. 338-381). New York: Praeger.

Kyllonen, P. C. (1987). *Componential analysis of semantic matching*. Unpublished manuscript, University of Georgia, Athens.

Kyllonen, P. C., & Shute, V. J. (in press). Learning indicators from a taxonomy of learning skills. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences*. New York: Freeman.

Kyllonen, P. C., Soule, C., & Stephens, D. (1988). *The role of working memory and general problem-solving skill in acquiring computer programming skill*. Unpublished manuscript.

Kyllonen, P. C., Stephens, D., & Woltz, D. J. (1988). *The role of working memory and accretive learning processes in learning logic gates*. Unpublished manuscript.

Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1985). *The speed-level problem reconsidered*. Manuscript submitted for publication.

Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1988). *Knowledge and processing speed as determinants of associative learning* (AFHRL-TP-87-68). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 4*, 477-494.

Learning Research and Development Center (1987). *Research in Intelligent CAI at the Learning Research and Development Center of the University of Pittsburgh*. Pittsburgh, PA: University of Pittsburgh, LRDC.

Lesgold, A. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory: Essays in honor of Gordon Bower* (pp. 31-64). San Francisco: Freeman.

Lesgold, A., Bonar, J., & Ivill, J. (1987, March). *Toward intelligent systems for testing* (Tech. Rep. No. LSP-1). Pittsburgh, PA: University of Pittsburgh, LRDC.

Marshalek, B. (1981, May). *Trait and process aspects of vocabulary knowledge and verbal ability* (Tech. Rep. No. 15). Stanford, CA: Stanford University, School of Education, Aptitude Research Project.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 61-88). Norwood, NJ: Ablex.

Posner, M. I., & McLeod, P. (1982). Information processing models--In search of elementary operations. *Annual Review of Psychology, 33*, 477-514.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190-214.

Rose, A. M., & Fernandez, K. (1977). *An information processing approach to performance assessment, I. Experimental investigation of an information processing performance battery* (Tech. Rep. No. 1). Washington, DC: American Institutes for Research.

Schank, R. C. (1980). How much intelligence is there in artificial intelligence? *Intelligence*, *4*, 1-14.

Schoenfeld, A. H. (1979). Explicit heuristic training as a variable in problem solving performance. *Journal for Research in Mathematics Education*, *10*, 173-187.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, *8*, 165-190.

Snow, R. E. (1979). Theory and method for research on aptitude processes. In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 105-137). Norwood, NJ: Ablex.

Stephens, D. L. (1987). *Use of cognitive structure in predicting test achievement and ideational creativity in biology students*. Unpublished master's thesis, University of Georgia, Department of Educational Psychology, Athens, GA.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1981a). Intelligence and nonentrenchment. *Journal of Educational Psychology*, *73*, 1-16.

Sternberg, R. J. (1981b). Testing and cognitive psychology. *American Psychologist*, *36*, 1181-1189.

Sternberg, R. J., & Powell, J. (1983). Comprehending verbal comprehension. *American Psychologist*, *38*, 878-893.

Tirre, W.C. & Elliott, L.R. (1987, December). *Development and validation of an experimental battery to assess the components of text comprehension*. Paper presented at the Annual Meeting of the National Reading Conference, St. Petersburg Beach, Florida.

Tirre, W. C., Royer, J. M., Greene, B. A., & Sinatra, G. M. (1987, April). *Assessing on-line comprehension in a computer based instruction environment*. Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*, 128-134.

Wilson, K. E. (1976). *The GRE Technical Manual*. Princeton, NJ: Educational Testing Service.

Woltz, D. J. (1987). *Two constructs of working memory capacity: Attention and activation.* Manuscript submitted for publication.

Woltz, D. J., & Christal, R. E. (1985, April). *Working memory*. Paper presented at the Western Psychological Association Annual Convention, San Jose, CA.