3. What Cognitive Psychology Can (and Can not) Do for Test Development

Robert J. Sternberg
Yale University, robert.sternberg@okstate.edu

Follow this and additional works at: http://digitalcommons.unl.edu/burostestingissues

Part of the Educational Assessment, Evaluation, and Research Commons

http://digitalcommons.unl.edu/burostestingissues/5

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Social and Technical Issues in Testing: Implications for Test Construction and Usage by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Whenever research is launched under a new paradigm for studying an old set of mental phenomena, researchers joining the new armada of explorers hope, at best, to discover new uncharted mental territories and, at worst, to provide new mental maps of previously charted territories that amend errors of the old maps. This has been, I believe, the experience of cognitive psychologists studying mental abilities. Although they may not have revolutionized our map of the mind (yet), neither have they left the old maps standing. What is critical is that at least the flaws and incompletenesses of the new methods are different from those of the old. One can therefore be provided with some new insights about the mental phenomena being studied. Consider an analogy to polar and Cartesian coordinates: Each provides a different and useful view of a world that is not quite so simple as either coordinate system would have us believe. Seeing the mental world in two ways can tell us more than seeing it in just one way. In the language of Garner, Hake, and Eriksen (1956), we have provided “converging operations” to view a unitary phenomenon.

I have divided my analysis of the contribution of cognitive psychology to test construction into four main parts dealing, respectively, with the contributions of cognitive psychology to: (1) content for construction of tests; (2) validation of tests; (3) scoring and interpretation of tests; and (4) modification of tests. Before discussing these contributions, however, let me say just what are the characteristics that define “cognitive psychology” and what psychologists do in the cognitive–psychological investigation of intelligence.
WHAT IS COGNITIVE PSYCHOLOGY?

Cognitive psychology is the study of the mind in terms of the mental (cognitive) representations and processes that underlie observable behavior. In particular, I find that cognitive researchers tend to address five main questions:

1. What are the mental processes that constitute intelligent task performance?
2. How rapidly and accurately are these processes performed?
3. Into what strategies for task performance do these mental processes combine?
4. Upon what forms of mental representation do these processes and strategies act?
5. What is the knowledge base that is organized into these forms of representation, and how does it affect and become affected by the processes, strategies, and representations that individuals use?

These questions have been asked of performance on a rather wide range of cognitive tasks.

Cognitive Versus Psychometric Approaches

The cognitive approach is often contrasted with the psychometric one, perhaps because historically it has seemed easy enough to separate the psychometricians of a given time from the experimental psychologists of that time (many of whom now call themselves “cognitive” psychologists). However, I think it worth mentioning and even emphasizing that the distinction has never been as clear as Cronbach’s (1957) paper on “the two disciplines of scientific psychology” or as the conventional wisdom (which may in part be based upon Cronbach’s paper) might have one believe. Many of the great experimental or “cognitive” psychologists in the history of psychology have also been psychometricians and vice versa, and it often seems almost arbitrary to identify a given individual as one or the other. Consider some examples.

Sir Francis Galton invented the correlational method and yet was an avid experimentalist. Alfred Binet invented the prototype for the most widely used psychometric intelligence test, and yet a close reading of his writings will show his often neglected theorizing to be as “cognitive” as any we find today. Charles Spearman invented factor analysis, and yet his 1923 treatise on the “principles of cognition” was a cognitive monograph and the basis for much cognitive theorizing today, particularly in the domain of inductive reasoning. Edward Thorndike is most well known for his experimental work in animal learning and yet was the author of a major book on the subject of psychometrically measured intelligence. Clark Hull, another famous learning theorist, wrote his first book on the subject of aptitude testing. Louis Thurstone, a psychometrician if ever there was one, advocated factorial methods as preliminaries to experimental ones, not
as replacements for them. Finally, J. P. Guilford, clearly identified as a psychometrician, has also proposed a theory of intelligence in which one of the three facets describes the processes of intelligence.

The list could go on and on, but I think the point by now should be clear: Even before Cronbach’s (1957) paper, there was already a substantial connection between psychometric and cognitive lines of endeavor. I do not believe they were ever quite so separate as Cronbach’s paper suggested, and for all the overwhelming positive contribution the paper clearly made, it may have served the slightly negative function of tending to underscore the points of friction rather than the points of smooth contact. Today, I believe (in part because of the positive contribution of Cronbach’s paper) that the distinction between psychometricians and cognitive psychologists, at least in the research domain of intelligence, is fuzzier than ever. People like Jack Carroll, Earl Hunt, Robert Glaser, Susan Whitely, Richard Snow, and myself, among others, could perhaps as easily be classified as falling into one camp as into the other.

If the line between psychometrics and cognitive psychology is so unclear, just what is it that is distinctive about the cognitive approach? Certainly it is not just the questions asked, in that the questions listed earlier would also be quite relevant to the interest of many psychometricians theorizing about intelligence. Nor is it, really, the emphasis of the questions upon stimulus rather than subject variation. Psychometric methods, like factor analysis, really can be used to study either source of variation (although they are most commonly used to study subject variation), and experimental methods can also be used to study either stimulus or subject variation (although they are most commonly used to study stimulus variation).

The critical difference, I believe, is a sociological one and resides primarily (but not exclusively) in the professional identification of the investigator and of the methods he or she uses. A number of contemporary investigators, including this author, use multiple regression in modeling of test performance; for whatever reason, this methodology today seems more to belong to the “cognitive camp,” despite the fact that multiple regression can certainly be and has been viewed as a psychometric method. Susan Whitely does a highly similar (although by no means identical) kind of modeling using latent-trait analysis and tends to be viewed more in the psychometric camp. Users of exploratory factor analysis, like Raymond Cattell and John Horn, tend to be identified with the psychometric camp, whereas users of confirmatory factor analysis, like Carl and John Frederiksen, tend more to be identified with the experimental camp. The lines between camps are certainly not clearly drawn, although they can be inferred to some extent by the conventions one attends and by the journals in which one publishes as well as by the methods one uses. The rationale for placing someone in one or the other camp is certainly not clear-cut.

I have tried in several ways to make the basic point, one that I have come to believe only recently. This point is that the lines that have been drawn between
the psychometric and cognitive approaches to intelligence are often arbitrary and even capricious. If I speak in this chapter of the contribution of cognitive psychology to the psychometric tradition of mental test development, I am speaking of boundaries between traditions that I think have much more to do with the sociology of science than with its substantive concerns.

Cognitive–Psychological Approaches to the Study of Mental Abilities

How do cognitive psychologists go about studying mental abilities? Cognitive psychologists are highly similar in their emphasis on intensive task analysis. The idea is to take performance on a single task and then to study it in great depth. One then constructs an information-processing model of performance in the given task, a model that specifies in considerable detail just how subjects solve the task. Only after the task has been intensively analyzed is an attempt made to generalize the results of the task analysis to related tasks as well.

One can carve up the field in many different ways, as people in fact have (Pellegrino & Glaser, 1979; Sternberg, 1977, 1981c). I have loosely classified these different approaches into four different categories, but it should be understood that these categories are neither mutually exclusive nor exhaustive with respect to current research approaches in cognitive psychology. I now briefly describe what each approach is, what its goals are, what kinds of research it has generated, and what its implications for test construction are.

Cognitive Correlates. In this approach to understanding mental abilities, subjects are tested in their ability to perform tasks that contemporary cognitive psychologists believe measure basic human information-processing abilities. (Information processing is generally defined as the sequence of mental operations and their products involved in performing a cognitive task.) Such tasks include, among others, the Posner and Mitchell (1967) letter-matching task, in which subjects are asked to state as quickly as possible whether the letters in a pair such as “A a” constitute a physical match (which they don’t) or (in another condition) a name match (which they do), and the S. Sternberg (1969) memory-scanning task, in which subjects are asked to state as quickly as possible whether a target digit or letter, such as 5, appeared in a previously memorized set of digits or letters, such as 3 6 5 2. Individuals are usually tested either via tachistoscope (a machine that provides rapid stimulus exposures) or via a computer terminal, with the principal dependent measure of interest being response time.

The proximal goal in this research is to estimate parameters (characteristic quantities) representing the durations of performance for the information-processing components constituting each task and then to investigate the extent to which these components correlate across subjects with each other and with scores on measures commonly believed to assess intelligence (e.g., Raven’s Progres-
Most commonly, correlations between parameter estimates and measured intelligence are statistically significant but moderately low—usually around .3 (Hunt, Frost, & Lunneborg, 1973; Hunt, Lunneborg, & Lewis, 1975). The distal goal of cognitive-correlates research is to integrate individual-differences research and mainstream cognitive—psychological research—in particular, by providing a theoretical grounding from cognitive psychology for differential research (Hunt et al., 1973). Thus, instead of trying to draw theoretical conclusions by correlating scores on one empirically derived test (e.g., reasoning) with scores on another empirically derived test (e.g., vocabulary), as differential researchers have done, cognitive-correlates researchers draw theoretical conclusions by correlating scores on an empirically derived test with parameters generated by a cognitive model of some aspect of mental functioning (e.g., memory scanning).

Cognitive-correlates researchers would be most likely to supplement psychometric tests with information-processing tests based on standard laboratory information-processing tasks such as the memory-scanning and letter-matching tests mentioned earlier. Rose (1978) has actually constructed and tested an information-processing assessment battery based on standard laboratory tasks. Using this battery, one can isolate latency scores for a variety of different information-processing components. Rose's battery is an impressive one, although correlations across tasks are relatively low, and correlations of the information-processing tasks and parameters with psychometric tests or various types of real-world performance have yet to be reported.

**Cognitive Components.** In this approach to understanding mental abilities, subjects are tested for their ability to perform tasks of the kinds actually found on standard psychometric tests of mental abilities—for example, analogies, series completions, mental rotations, and syllogisms. Subjects are usually tested via a tachistoscope or a computer terminal, and response time is usually the principal dependent variable, with error rate and pattern of response choices as secondary dependent variables. These latter dependent variables are of more interest in this approach than in the cognitive-correlates approach because the tasks tend to be more difficult and thus more susceptible to erroneous responses.

The proximal goal in this research is first to formulate a model of information processing in performance on IQ-test types of tasks; second, to test the model at the same time as parameters for the model are estimated; and, finally, to investigate the extent to which these components correlate across subjects with each other and with scores on standard psychometric tests. Because the tasks that are analyzed are usually taken directly from IQ tests, or else are very similar to tasks found on IQ tests, the major issue in this kind of research is not whether there is any correlation at all between cognitive task and psychometric test scores. Rather, the issue is one of isolating the locus or loci of the correlation that is obtained. One seeks to discover what components of information processing in
task performance are the critical ones from the standpoint of the theory of intelligence.

Cognitive-components researchers would be most likely to supplement psychometric tests with information-processing tests based on the psychometric ones, but with test items administered in a computer-controlled setting that would enable the examiner to decompose test performance into its information-processing constituents. An information-processing analysis of a subject’s inductive reasoning performance, for example, would assess skills such as the individual’s ability: (1) to encode stimuli; (2) to infer relations between stimulus terms; (3) to map higher-order relations between relations; (4) to apply previously inferred relations to new settings; (5) to compare alternative answer options in terms of their similarities and differences; (6) to justify one answer as preferred but not necessarily ideal; (7) to respond; (8) to combine these components into a strategy that results in efficient item solution; and (9) to represent information in a way that facilitates operations on the data base stored in long-term memory (Sternberg & Gardner, 1982).

Cognitive Training. This approach to understanding mental abilities can be used in conjunction with either the cognitive-correlates approach or the cognitive-components approach, or in conjunction with some other approach. The essence of this approach is aptly described by Campione, Brown, and Ferrara (1982). Essentially, the approach seeks to infer the identities of cognitive processes through an analysis of effects of training. The cognitive-training approach has been used widely in a variety of domains. For example, Belmont and Butterfield (1971), Borkowski and Wanschura (1974), and Campione and Brown (1978) have used the approach in investigations of learning and memory. Feuerstein (1979), Holzman, Glaser, and Pellegrino (1976), and Linn (1973) have used it in investigations of reasoning and problem solving. One conclusion has emerged with striking regularity in many studies by many different investigators: To attain both durability and generalizability of training, it seems to be necessary to train both at the level of metacomponents (or executive processes) and at the level of performance components (or lower-order processes used to carry out the orders of the executive processes—see, e.g., Belmont, Butterfield, & Ferretti, 1982; Feuerstein, 1979, 1980).

Cognitive-training researchers might follow any of a number of paths to testing depending on their choice of what to train and how to train it. One of the more interesting approaches to testing among such investigators is that of Feuerstein (1979), who has devised a “learning potential assessment device” that he believes measures cognitive modifiability, or what Vygotsky (1978) referred to as the “zone of potential development.” Modifiability is assessed by giving examinees guided instruction in solving problems that the examinees are initially unable to solve and by evaluating the examinees’ ability to profit from instruction.
Cognitive Contents. Recently, a new approach to research has emerged on the cognitive–psychological scene that has yet to be applied directly to the study of mental abilities but that seems to provide a good entree into such research. The approach seeks to compare the performances of experts and novices in complex tasks such as the solution of physics problems (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980, 1980b), the selection of moves and strategies in chess and other games (Chase & Simon, 1973; DeGroot, 1965; Reitman, 1976), and the acquisition of domain-related information by groups of people at different levels of expertise (Chiesi, Spilich, & Voss, 1979; Spilich, Vesonder, Chiesi, & Voss, 1979). Research on expert–novice differences in a variety of task domains suggests that the way information is stored in and retrieved from long-term memory can largely account for the substantial differences in performance between experts and novices. This view would suggest that a possible locus of differences between more and less mentally able people is in their ability to organize information in long-term memory in a way that makes it readily accessible for a variety of purposes (Egan & Greeno, 1973). Presumably, information stored in such a flexible way is maximally available for transfer from old to new problem situations.

Because the cognitive-contents approach has not yet been directly applied to the investigation of differences in mental abilities, it is impossible to evaluate its utility for purposes of such investigation. But the approach seems to supply a valuable new inroad for mental-abilities research, and I expect it will be only a matter of time before it is used for this purpose.

Cognitive-contents researchers might supplement psychometric tests with complex-learning or problem-solving tasks that elicit an examinee’s knowledge base and the way in which knowledge is mentally represented. Such researchers would be particularly interested in the features of problems to which examinees attend. It has been found, for example, that less skilled physics problem solvers tend to pay more attention to surface features of physics problems, whereas more skilled problem solvers tend to pay more attention to deep structural features (Chi et al., 1981, 1982). Cognitive-contents tests might also supplement cognitive-components tests, with the former assessing knowledge deficiencies and the latter assessing processing deficiencies.

Cognitive psychologists studying mental abilities differ markedly in the tasks they have chosen to study, in the dependent variables they use to study these tasks, in the kinds of theories that motivate their research, and in their concern with individual differences. Because I have more to offer later about differences in task content, I pass over this source of differences now. The kinds of dependent variables that cognitive psychologists use include reaction time, percentage correct, breakdown of response choices, protocol analysis, and output of computer simulations. Although choice of one (or more) dependent variable(s) may not seem like a major issue to many psychologists, the history of cognitive psychology up to the present has been marked by active (and at times explosive)
debates regarding the relative merits of various dependent measures (see, e.g., Ericsson & Simon, 1980; Nisbett & Wilson, 1977; Pachella, 1974; Sternberg, 1977; S. Sternberg, 1969). Cognitive psychologists also differ in the scope of the theories that motivate their research, with scope ranging from quite narrow (Egan, 1976) to very broad (Anderson, 1976). The optimal scope of a motivating theory of intelligence has also been a subject of intense debate among cognitive psychologists (Anderson, 1976; Sternberg & Davidson, 1982). Finally, cognitive psychologists differ greatly in their concern with individual differences. In the past decade, cognitive psychologists have progressed from little or no concern to an increasingly broad concern with this issue.

COGNITIVE PSYCHOLOGY AND TEST CONSTRUCTION

Cognitive psychologists studying mental abilities have investigated a wide range of tasks, some of which have been used in test construction. The tasks they have investigated differ in a multitude of ways, but it is convenient and, I believe, accurate to array them along a single dimension of task complexity, from simple and choice reaction time at one extreme to complex logic and mathematics at the other extreme. The tasks along this continuum differ in the apparent “level” of mental processing required.

At the simple end of the continuum, Furneaux (1956), Jensen (1979), and Lunneborg (1977) have used simple and choice reaction time tasks to test the hypothesis that individual differences in mental ability can be understood largely in terms of individual differences in sheer speed of mental functioning. Hunt (1978; Hunt et al., 1975) has studied mental speed as well but at a somewhat higher level of processing. He has suggested that individual differences in mental abilities, especially verbal ones, can be understood in terms of differences in people’s speed of access to lexical information in long-term memory. In sharp contrast to Furneaux, Jensen, and Lunneborg, he has preferred to hold constant simple or choice reaction time divorced from lexical access so as not to confound his measurement of access speed. Pellegrino and Glaser (1979, 1980), Snow (1979), and I (Sternberg, 1977, 1980c, 1981c), among others, have claimed that the level of mental processing studied by Hunt and his colleagues is still low and have preferred instead to study performance on tasks at a level of complexity equal to that of intelligence-test items. Like those cited earlier, these researchers have emphasized speed of processing, but particularly speed in solving relatively complex tasks such as analogies and syllogisms. Finally, investigators such as Greeno (1980); Chi et al. (1982); and Larkin (1981) have suggested, if only by implication, that even the intelligence-test items are at too low a level of processing and have studied instead performance in very complex mathematics and physics problems.

In sum, the range of tasks studied by cognitive psychologists investigating intelligence is at least as broad as that studied by psychometricians. Indeed, the
range in levels of complexity is probably greater: Whereas most psychometrists seem to have resolved the Galton–Binet dispute regarding test content to their satisfaction, cognitive psychologists seem not to have done so.

Even if cognitive psychologists did display more agreement regarding the kinds of task performance that should be studied under the rubric of intelligence, it is not at all clear that they would have much to contribute to psychometricians by way of useful feedback regarding test content, because when cognitive psychologists have used reference measures at all for external criteria for their tasks and theories, they have used intelligence tests and subtests rather than the behaviors these tests were themselves intended to predict (such as school grades and job success). Their use of psychometric tests as (obviously proximal) criteria for their own tasks and tests has made it impossible to use their data to modify the tests. One can use the criteria to suggest changes in the predictor but not the other way around!

My message regarding the contribution of cognitive psychology to selection of test content is not a wholly pessimistic one, however. Some recent cognitive research has suggested promising lines of endeavor that I believe are now ready for at least pilot attempts in psychometric tests. I think three suggestions are clearly forthcoming, albeit from experiments using IQ-test items as criteria.

First, there is good evidence that performance on the Clark and Chase (1972) sentence–picture verification task, which requires the examinee rapidly to indicate whether a sentence representation (such as "The star is above the plus.") agrees with a pictorial representation (such as \( \star \)), can provide a quick estimate of a person's general level of intelligence (Hunt et al., 1975) and even, possibly, of their proclivity for verbal versus spatial strategies for problem solving (MacLeod, Hunt, & Mathews, 1978; Mathews, Hunt, & MacLeod, 1980). The task is easy to administer and usable for examinees over a wide range of ages, ability levels, and mental conditions.

Second, there is strong evidence to suggest value in measuring fluid intelligence by using novel tasks employing novel kinds of concepts (Snow, 1980; Sternberg, 1981a). The important thing appears to be not the particular task or concepts used but their relative novelty for the examinees performing them. By novelty, I refer not only to a difference in content but to a difference in kind from conventional kinds of test items.

Third, substantial evidence has now been accumulated for the considerable value in measuring crystallized, or verbal, intelligence of a task requiring examinees to learn and then define previously unfamiliar words presented in natural written contexts (Sternberg, Powell, & Kaye, 1982). Such a task appears to tap at least one major aspect of the antecedents of developed individual differences in verbal skills and knowledge.

To conclude this section, cognitive psychology has probably not been at its best in suggesting to test developers the kinds of content they might profitably use in test construction. There is almost no resolution among cognitive psychologists as to what kinds of test contents best measure intellectual functioning, and
experiments have not been designed in ways that would be particularly informative with regard to suggested content even if cognitive psychologists could agree as to what kinds of contents to employ. Nevertheless, a few suggestions have emerged from cognitive research regarding several kinds of contents that might be beneficially employed in future testing.

COGNITIVE PSYCHOLOGY AND TEST VALIDATION

Whereas cognitive psychology has probably made its weakest contribution to test development in the realm of test content, it has probably made its strongest contribution in the realm of test validation and, in particular, construct validation. There is perhaps some irony in the fact that the paradigm that was perhaps hoped by some to provide a replacement for the psychometric paradigm has instead provided converging evidence to support its major findings. Let me elaborate.

Whereas psychometricians have generally attempted to understand mental abilities through the construct of the "factor," cognitive psychologists have generally attempted to understand mental abilities through the construct of the "process" and, to a lesser extent, the "mental representation" of information. Through successive refinements, cognitive psychologists have developed techniques that seem to be quite successful in the isolation of mental processes (Ericsson & Simon, 1980; Newell & Simon, 1972; Pachella, 1974; Siegler, 1976; Sternberg, 1977; S. Sternberg, 1969). A few of the cognitive psychologists, such as Carroll (1976, 1981) and myself (Sternberg, 1980a), have explicitly addressed the question of the extent to which the structural factors of psychometricians deal, at some deep level, with the same latent abilities as the processes of cognitive psychologists. We have concluded that both sets of investigators are, in fact, looking at the same underlying entities, albeit in different ways.

I disagree with Carroll’s (1981) position that factors are in some interesting sense more "basic" than are processes, and I also disagree with my own earlier position (Sternberg, 1977) that processes are in some interesting sense more basic than factors. So far as I can tell, there exists no empirical means to determine which is more basic, nor is it even clear what, conceptually, "more basic" means. If there is some basic molar unit in terms of which mental abilities are organized, we probably do not know what it is; even more discouraging, perhaps, we wouldn’t know we knew if we did, in fact, know. At this point, therefore, I regard arguments regarding basic-level mental units as nonfruitful and believe we should probably be quite pleased that constructs from the psychometric and cognitive, as well as from other approaches (Sternberg, 1981b), have converged as well as they have in suggesting how mental abilities might be organized.
What, exactly, are these points of convergence? I believe there are three main ones.

First, there appears to be some (if you wish, higher-order) general factor or source of individual differences that is common to performance on a strikingly wide range of cognitive tasks (Holzinger, 1938; Jensen, 1980; Spearman, 1927; Thurstone, 1938; Vernon, 1971). Individual differences in this general ability, or \( g \), appear to derive in large part from differences in the functioning of (higher-order) executive processes—such as solution planning, monitoring, and control—that regulate most mental functions (Butterfield & Belmont, 1977; Campione & Brown, 1978; Snow, 1979; Sternberg, 1979).

Second, there appear to be at least two, and possibly several more, broad constellations of skills that operate in fairly broad ranges of tasks but not, by any means, in all tasks. The two most prominent constellations, which have been referred to by many names but here will be referred to by the names of “fluid” and “crystallized” abilities (Horn & Cattell, 1966), encompass reasoning kinds of tasks on the one hand and verbal kinds of tasks on the other (Cattell, 1971; Horn, 1968; Vernon, 1971). Individual differences in these abilities appear to be traceable to present and past functioning of lower-order performance and learning processes, as well as to the interactions of these processes with the higher-order executive ones (Sternberg, 1980c).

Finally, for however they may interact, it is important to separate speed from power aspects of performance (Carroll, 1981; Egan, 1976; Guyote & Sternberg, 1981; Sternberg, 1977, 1980b). Speed and power appear to be differentiable aspects of mental skill, and confounding them can lead to misleading or even downright incorrect conclusions (Sternberg, 1980b).

To conclude this section, I would argue that cognitive psychology has provided a valuable complementary way of investigating pretty much the same constructs psychometricians have been studying all along. The contribution of cognitive psychology goes beyond a merely salutary or congratulatory one. Cognitive psychology has provided insights into the processes underlying the products studied by psychometricians and has told us what happens in real time to generate these products. The process models of cognitive psychologists, and the theoretical and metatheoretical schemes underlying them, have provided important insights into mental abilities that previously had been lacking.

**COGNITIVE PSYCHOLOGY, TEST SCORING, AND TEST INTERPRETATION**

Using a cognitive approach, one would derive and interpret a set of test scores quite different from that derived and interpreted via a psychometric approach. The major difference in scoring would be the isolation, in cognitive analysis, of subscores based on processes rather than actual or alleged factors.
Consider, for example, the rather global construct of reasoning ability. It would not be at all surprising to discover that individuals believed, for one reason or another, to be of low intelligence score below the average on tests of reasoning ability. But exactly what does this tell us? Does it tell us what it is that leads to the subject's low intelligence? Does it tell us what kinds of interventions might be indicated to increase the individual's level of intellectual functioning? Does it even tell us that the individual is low in reasoning ability as opposed, say, to encoding the terms of the problem so that the reasoning operations can be performed? I would argue that the answer to each of these questions is negative; in short, that the low score in reasoning provides relatively little by way of diagnostic or prescriptive information.

A cognitive analysis of the bases of performance on one or more reasoning tests would seek to go “inside” the reasoning factor—to elicit for each individual a measure of performance on each of the processes theorized in combination to constitute reasoning performance. In my own theory of inductive reasoning, for example (Sternberg, 1977, 1980c), separate component scores might be estimated for processes such as encoding, inference, application, and response. Other cognitive theories would yield somewhat different sets of process scores, just as alternative factorial theories yield somewhat different sets of factor scores. At the very least, any of the “reasonable” theories of cognitive processing would permit a separation between performance on the encoding, pure reasoning, and response aspects of task performance.

These process scores not only permit a finer diagnosis of strengths and weaknesses in cognitive skills but permit as well the construction of a process-based training program. It is difficult to conceive of training something as ill-defined as “reasoning” but relatively easy to conceive of training a specific skill such as inferring relations. The relative ease of conceiving and actually preparing such a training program should not, however, desensitize one to the considerable difficulty that can be involved in instantiating transfer of training in the individuals exposed to the program of instruction.

The theoretical basis now exists not only for analyzing processing skills in reasoning tasks but for analyzing processing skills in other kinds of tasks as well, such as spatial, verbal, and numerical tasks. Yet, I do not recommend our actually implementing the theory in practice at this time. There are several reasons for my reluctance to what I view as premature implementation.

First, obtaining reliable estimates of process scores for individuals requires very lengthy testing, usually via a computer terminal or comparable device. Thus, the technology does not yet exist for implementing theory in an expeditious way. We need much more research aimed simply at enabling efficient measurement of process parameters of test performance.

Second, the differential validity of process scores in predicting interesting criterion performances has yet to be demonstrated. At present, such differential validity is available as a promissory note rather than as a demonstrated accomplishment.
Third, although we have the means for isolating lower-order processes of performance (i.e., those processes used in strategy implementation), we do not yet have adequate means for isolating higher-order executive processes (i.e., those processes used in strategy planning, monitoring, and control). Yet, these latter processes are the ones I believe most crucial to understanding the bases of individual differences in intelligence. Until we have a feasible technology for isolating these more interesting processes, I am reluctant to advocate rapid implementation of process analysis in mental-ability testing.

To conclude this section, I believe we now have a theoretical basis for the scoring and interpretation of ability tests but that the practical basis has lagged behind. In some ways, this situation is a welcome contrast to what has been the typical one in abilities research, where theory has tended to lag behind practice. I believe that process analysis will eventually become both feasible and desirable in the scoring and interpretation of ability tests; the time has not yet come.

COGNITIVE PSYCHOLOGY AND TEST MODIFICATION

What are the implications of the previous discussion for the modification of ability tests? The answer depends on the time frame into which one puts it. At present, I think they are modest. None of the cognitive research that has been done has come up with any alternative test that is clearly better than the best of the ability tests we now have. But there have been interlaced throughout this discussion a number of promising notes that I would like to summarize here, because I believe they will, eventually, result in test modification. First, with regard to test content, I feel the research to date suggests the importance of using measurements based upon performance on novel tasks comprising novel task content. Second, with regard to construct validation, I think cognitive research has shown that current tests can be understood in terms of their measurement of process constructs. Third, with regard to test scoring and interpretation, I believe it will eventually be possible to measure executive and performance processes in technically feasible ways and that such measurements will provide new bases for diagnosis and training that are currently unavailable. Finally, I feel that cognitive psychology will continue to provide a basis for the questioning of some of our assumptions regarding the nature of mental abilities and how they can be measured.

Let me give three specific examples of some dubious assumptions regarding the nature of mental abilities that are entrenched in mental testing, and let me show how these assumptions are being added into question by information-processing research.

_Dubious Assumption 1. To be smart is to be fast._ The assumption that "smart is fast" permeates our entire society. When we refer to someone as "quick," we are endowing them with one of the primary attributes of what we
perceive an intelligent person to be. Indeed, in a recent study of people’s conceptions of intelligence, when we asked people to list behaviors characteristic of intelligent persons, behaviors such as “learns rapidly,” “acts quickly,” “talks quickly,” and “makes judgments quickly” were commonly listed (Sternberg, Conway, Ketron, & Bernstein, 1981). It is not only the man in the street who believes that speed is associated with intellect: Several prominent contemporary theorists of intelligence base their theories in large part upon individual differences in the speed with which people process information (Hunt, 1978; Jensen, 1979).

The assumption that more intelligent people are rapid information processors also underlies the overwhelming majority of tests, including creativity as well as intelligence tests. It is rare to find a test that is not timed or a timed test that virtually all examinees are able to finish by working at a comfortable rate of problem solving. I would argue that this assumption is a gross overgeneralization: It is true for some people and for some mental operations but not for all people or all mental operations. Blind, across-the-board acceptance of the assumption is not only unjustified—it is wrong.

Almost everyone knows people who, although often slow in performing tasks, perform the tasks at a superior level of accomplishment. Moreover, we all know that snap judgments are often poor ones. Indeed, in our study of people’s conceptions of intelligence, “does not make snap judgments” was listed as an important attribute of intelligent performance. Evidence for the dubiousness of the “smart is fast” assumption extends, however, beyond intuition and everyday observation. A number of findings from carefully conducted psychological research undermine the validity of assumption. I will cite four such findings, which are only examples from a wider literature on the subject.

First, it is well known that, in general, a reflective rather than an impulsive style in problem solving tends to be associated with more intelligent problem-solving performance (see Baron, 1981, 1982, for reviews of this literature). Jumping into problems without adequate reflection is likely to lead to false starts and erroneous conclusions. Yet, timed tests often force the examinee to solve problems impulsively. It is often claimed that the strict timing of such tests merely mirrors the requirements of our highly pressured and productive society. But ask yourself how many significant problems you encounter in your work or personal life that allow no more than the 15 to 60 seconds allowed for a typical test problem on a standardized test; you will probably be hard pressed to think of any such problems.

Second, in a study of the role of planning behavior in problem solving, it has been found that more intelligent persons tend to spend relatively more time than do less intelligent persons on global (higher-order, up-front) planning and relatively less time on local (problem-specific, lower-level) planning. In contrast, less intelligent persons show the reverse pattern, emphasizing local rather than global planning (relative to the more intelligent persons) (Sternberg, 1981a). The
point is that what matters is not total time spent but distribution of this time across the various kinds of planning one can do.

Third, in studies of reasoning behavior in children and adults, it has been found that although greater intelligence is associated with more rapid execution of most components of information processing, problem encoding is a notable exception to this trend. The more intelligent individuals tend to spend relatively more time encoding the terms of the problem, presumably to facilitate subsequent operations on these encodings (Mulholland, Pellegrino, & Glaser, 1980; Sternberg, 1977; Sternberg & Rifkin, 1979). Similar outcomes have been observed in comparisons of expert versus novice problem solvers confronted with difficult physics problems (Chi et al., 1982).

Finally, in a study of people’s performance in solving insight problems (arithmetical and logical problems whose difficulty resided in the need for a nonobvious insight for problem solution rather than in the need for arithmetical or logical knowledge), a correlation of .75 was found between the amount of time people spent on the problems and measured IQ. The correlation between time spent and score on the insight problems was .62 (Sternberg & Davidson, 1982).

Note that, in these problems, individuals were free to spend as long as they liked solving the problems. Persistence and involvement in the problems was highly correlated with success in solution: The more able individuals did not give up; nor did they fall for the obvious, but often incorrect, solutions.

The point of these examples is simple: Sometimes speed is desirable; sometimes it is not. Whether it is desirable or not depends on the task, the particular components of information processing involved in solution of the task, and, most likely, the person’s style of problem solving. Blind imposition of a strict time limit for a test, or even a not-so-strict one, is theoretically indefensible and practically self-defeating.

Dubious Assumption 2. Intelligence is last year’s achievement. At first glance, this would appear to be an assumption few people would accept. Indeed, doesn’t almost everyone make a clear distinction between intelligence and achievement? But if one examines the content of the major intelligence tests, one will find that they measure intelligence as last year’s (or the year before’s, or the year before that’s) achievement. What is an intelligence test for children of a given age would be an achievement test for children a few years younger. In some test items, like vocabulary, the achievement component is obvious. In others, it is more disguised, for example, verbal analogies or arithmetic problems. But virtually all tests commonly used for the assessment of intelligence place heavy achievement demands on the students tested.

The achievement-testing orientation exhibited in intelligence tests may be acceptable and even appropriate when the tests are administered to children who have had fully adequate educational opportunities in reasonably adequate social and emotional environments. But for children whose environments have been
characterized by deprivation of one kind or another, the orientation may lead to invalid test results. There is no fully adequate solution to the problem of identification of the gifted among such youngsters, especially if the youngsters will have to function in the normal sociocultural milieu. A common solution to the problem, exclusive use of nonverbal tests, is almost certainly an inadequate solution: First, one is measuring only a subset of important intellectual skills; second, and perhaps more importantly, nonverbal tests actually show, on the average, greater differences in scores across sociocultural groups than do verbal ones (Jensen, 1980; Lesser, Fifer, & Clark, 1965). An alternative solution to the problem is to ask what abilities one is really interested in measuring by the achievement-saturated tests and then to attempt to measure these abilities more directly and in ways that reduce the achievement load. This is the path we have followed. Consider two examples.

Consider first one of the most common types of items on intelligence tests—vocabulary. It is well known that vocabulary is one of the best predictors, if not the best single predictor, of overall IQ score (Jensen, 1980; Matarazzo, 1972). Yet, few tests have higher achievement load than does vocabulary. Can one measure the latent ability tapped by vocabulary tests without presenting children with what is essentially an achievement test? I believe one can.

There is reason to believe that vocabulary is such a good measure of intelligence because it measures, albeit indirectly, children's ability to acquire information in context (Jensen, 1980; Sternberg, Powell, & Kaye, 1982, 1983; Werner & Kaplan, 1952). Most vocabulary is learned in everyday contexts rather than through direct instruction. Thus, new words are usually encountered for the first time (and subsequently) in textbooks, novels, newspapers, lectures, and the like. More intelligent people are better able to use surrounding context to figure out the words' meanings. As the years go by, the better decontextualizers acquire the larger vocabularies. Because so much of one's learning (including learning other than vocabulary) is contextually determined, the ability to use context to add to one's knowledge base is an important skill in intelligent behavior. Is there any way of measuring this skill directly rather than relying on indirect measurement (vocabulary testing) that involves a heavy achievement load? We have attempted to measure this skill directly by presenting children with paragraphs written at a level well below their grade level. Embedded in the paragraphs are one or more unknown words. The children's task is to use the surrounding context to figure out the meanings of the unknown words. Note that, in this testing paradigm, differential effects of past achievement are reduced by using reading passages that are easy for everyone but target vocabulary words that are unknown to everyone. We have found that quality of children's definitions of the unknown words is highly correlated with overall verbal intelligence, reading comprehension, and vocabulary test scores (about .6 in each case). Thus, one can measure an important aspect of intelligence directly and without heavy reliance on achievement rather than indirectly and with heavy reliance on past achievement.
Consider second another common type of intelligence test—arithmetic word problems (and at higher levels, algebra and geometry word problems as well). Again, performance on such problems is heavily dependent on one’s mathematical achievements and, indeed, opportunities. Can one measure the main skills tapped by such tests without creating what is essentially an achievement test? We believe we have done so through the insight problems mentioned earlier. Consider two typical examples of such problems:

1. If you have black socks and brown socks in your drawer, mixed in the ratio of 4 to 5, how many socks will you have to take out to make sure of having a pair the same color?
2. Water lilies double in area every 24 hours. At the beginning of the summer there is one water lily on a lake. It takes 60 days for the lake to become covered with water lilies. On what day is the lake half-covered?

Solutions of problems such as these requires a fair amount of insight but very little in the way of prior mathematical knowledge. In most problems such as these, a common element in a successful solution is selective encoding—knowing what elements of the problem are relevant to solution and what aspects are irrelevant. Performance on such problems is correlated .66 with IQ. Thus, it is possible to use word problems that are good measures of intelligence but that require very little in the way of prior arithmetical knowledge (Sternberg & Davidson, 1982). Moreover, it is unnecessary to time problem administration. As mentioned earlier, higher performance is associated with more, not less, time spent on the problems.

To summarize: We need not measure intelligence as last year’s achievement. It is probably impossible to rid intelligence tests of achievement load entirely. Indeed, it may not even be desirable to do so. But the load can be substantially reduced by asking oneself what intellectual skills one wishes to measure and then by seeking to measure these directly through the use of items that tap the skills rather than their by-products.

_Dubious Assumption 3. Testing needs to be conducted in a stressful, anxiety-provoking situation._ Few situations in life are as stressful as the situation confronting the examinee about to receive (and then receiving) a standardized test. Most examinees know that the results of the test are crucial for the examinees’ future and that 1 to 3 hours of testing may have more impact on the future than years of school performance. The anxiety generated by the testing situation may have little or no effect on some examinees and even a beneficial effect on other examinees. But there is a substantial proportion of examinees—the test anxious—whose anxiety will cripple their test performance, possibly severely. Moreover, because the anxiety will be common to standardized testing situations (although often not to other testing situations), the error in measurement resulting from a single testing situation will be compounded by error in measurement in
other testing situations. With repeated low scores, a bright but test-anxious individual may truly appear to be stupid. What is needed is some kind of standardized assessment device that is fair to the test anxious, as well as to others, and that does not impose a differential penalty on individuals as a function of a form of state anxiety that may have no counterpart in situations other than that of standardized testing. I believe that we have at least two promising leads in this direction.

The first lead is testing based on the notion of intelligence as in part a function of a person's ability to profit from incomplete instruction (Resnick & Glaser, 1976). A measure of this ability is now provided by Feuerstein's (1979) Learning Potential Assessment Device (LPAD), which although originally proposed as an assessment device for retarded performers, can be used for performers at varying levels of performance, including advanced ones. The device involves administration of problems with graded instruction. The amount of instruction given depends on the examinee's needs. Moreover, the test is administered in a supportive, cooperative atmosphere, where the examiner is actually helping the examinee solve problems rather than impassively observing the examinee's success or failure. The examiner does everything he or she can do to allay anxiety (rather than to create it!). Feuerstein has found that children who are cowed by and unable to perform well on regular standardized tests can demonstrate high levels of performance on his test. Moreover, their performance outside the testing situation appears to be predicted better by the LPAD than by conventional intelligence tests (Feuerstein, 1979).

The second lead is based on the notion that intelligence can be measured with some accuracy by the degree of resemblance between a person's behavior and the behavior of the "ideally" intelligent individual (Neisser, 1979). Sternberg et al. (1981) had a group of individuals rate the extent to which each of 250 behaviors characterized their own behavioral repertoire. A second group of individuals rated the extent to which each of the 250 behaviors characterized the behavioral repertoire of an "ideally intelligent" person. The investigators then computed the correlation between each person's self-description and the description of the ideally intelligent person (as provided by the second group of individuals). The correlation provided a measure of degree of resemblance between a real individual and the ideally intelligent individual. The claim was that this degree of resemblance is itself a measure of intelligence. The facts bore out the claim: The correlation between the resemblance measure and scores on a standard IQ test was .52, confirming that the measure did provide an index of intelligence as it is often operationally defined. And doing self-ratings involved minimal stress.

The behaviors that were rated had previously been listed by entirely different individuals as characterizing either "intelligent" or "unintelligent" persons. The intelligent behaviors were shown (by factor analysis) to fall into three general classes: problem-solving ability (e.g., "reasons logically and well," "identifies connections among ideas," and "sees all aspects of a problem");
verbal ability (e.g., "speaks clearly and articulately," "is verbally fluent," and "reads with high comprehension"); and social competence (e.g., "accepts others for what they are," "admits mistakes," and "displays interest in the world at large"). (No attempt was made to classify the unintelligent behaviors, which were not the object of interest in the study.)

I would not propose the behavioral checklist, or the LPAD, for that matter, as replacements for standard intelligence tests. Certainly, there is not enough validity information yet to make such a proposition. But I think that they deserve to be considered as supplements to standard tests. They are much less stress provoking than standard intelligence tests and may well be more accurate, at least for individuals who fall to pieces when confronted with standardized tests. Persons who scored high on these new indices but low on conventional indices would merit further follow-up before writing them off as weak or even average performers. Such measures carry the potential of identifying gifted individuals who are being lost for no reason other than their high levels of test anxiety.

To conclude, although cognitive psychology has not yet provided (and may never provide) a basis for replacing existing psychometric tests, it has made and is continuing to make contributions that I believe will result in some important reconceptualizations of the nature of intelligence and its measurement. I think cognitive psychologists have made substantial progress toward this goal during the past decade, and I see no reason to believe that this progress will not continue.

REFERENCES


Sternberg, R. J. Factor theories of intelligence are all right almost. *Educational Researcher*, 1980, 9, 6–13, 18, (a)


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


