

Question: Do Standardized Tests Measure General Cognitive Skills? Answer: No

Standardized tests purport to measure the cognitive abilities required for a variety of academic and nonacademic tasks. Thus, test scores determine such important outcomes as whether students will be admitted into college, what grade they will receive, and the academic track where they will be placed.

Unfortunately, there have been few research attempts to validate the assumption that standardized tests do, in fact, measure general cognitive skills. Consequently, researchers at the Mid-continent Regional Educational Laboratory analyzed 6,942 items from the Stanford Achievement batteries (the Stanford Early School Achievement Battery, the Stanford Achievement Test, and the Stanford Test of Academic Skills) and the California Test of Basic Skills to identify the general cognitive abilities tested and study their relationship to student performance (Marzano and Jesse 1987).

General cognitive abilities were defined as those mental processes used in academic tasks that intersect more than one academic discipline. This definition stems from a widely accepted theory that the human mind stores information in two primary



Photograph by Sr. Mary Megan Dull

Performing well on standardized tests has little to do with students' thinking abilities and a great deal to do with how well they have learned factual knowledge represented in test items.

forms: (1) as factual or declarative knowledge, and (2) as process or procedural knowledge. Declarative knowledge is knowledge of *who*, *what*, *where* and *when* (e.g., knowledge of such facts as the nature and meaning of the number 3 and the nature and meaning of the term *equal*.) Procedural knowledge is knowledge of *how* (e.g., *how* to perform the process of dividing 3 into 4).

The example of procedural knowledge (dividing 3 into 4) is not a general process skill because it is specific to mathematics. However, such processes as comparing, contrasting, categorizing, and creating metaphors are general because they can be used in several content areas (e.g., you can create a metaphor to help understand a science problem, or to help understand a piece of literature, or to help understand a mathematics problem).

In all, the 6,942 items were analyzed for their inclusion of 22 general cognitive processes or thinking skills. These were drawn from the work of a number of thinking skills theorists who have identified and described those general cognitive processes that should be taught and reinforced in the classroom (Costa 1985, Ennis 1985, Marzano et al. 1987). Figure 1 contains

a brief description of the 22 general cognitive processes.

Two Findings

Two major findings emerged from the analysis of the Stanford and CTBS achievement batteries: (1) the test items included only 9 of the 22 general cognitive operations, and (2) the general cognitive operations required to answer the questions had very little to do with student achievement on those tests.

Relative to the first finding, the nine general cognitive operations identified within the 6,942 items analyzed were:

- comparing and contrasting
- inferring
- ordering
- reference
- retrieving
- representing
- summarizing
- transposing
- visual matching

Of these, comparing and contrasting and retrieving are necessary to answer virtually every item. That is, to answer every item analyzed, students need to retrieve information from long-term memory and compare and contrast

“Standardized tests in their present form are primarily measures of factual or declarative information.”

1. **Categorizing:** placing elements into superordinate and subordinate groups based on identified characteristics.
2. **Comparing and contrasting:** identifying similarities and dissimilarities among elements.
3. **Creating analogies:** identifying and creating sets of elements with similar relationships between the components within each set.
4. **Creating metaphors:** identifying or creating relationships between elements that are commonly considered unrelated.
5. **Dialectic thinking:** identifying and articulating a set of values contrary to your own for a particular set of information.
6. **Encoding:** storing information in long-term memory in such a way that it is easily retrieved.
7. **Establishing criteria:** setting and articulating standards for judging the logic or value of something.
8. **Extrapolating:** identifying how the basic theme in one piece of information is similar to and different from the basic theme in another piece of information.
9. **Identifying errors:** recognizing and articulating fallacies that result from unwarranted assumptions and fallacies that result from ambiguity.
10. **Identifying patterns and relationships:** identifying the component parts of a whole and articulating the relationships among the parts.
11. **Inferring:** deducing or inducing unstated information from stated information or filling in gaps in stated information using prior knowledge.
12. **Ordering:** identifying the attributes of elements in relative or absolute terms and then ranking or sequencing them according to those attributes.
13. **Predicting:** making a statement about the future and then investigating the outcome of the prediction as it relates to the original statement.
14. **Reference:** identifying explicit or implicit information as cued by syntax, pronouns, synonyms, or subordinate and superordinate terms.
15. **Restructuring:** reorganizing stated information into a new pattern or format that is different from that which is explicit in the information.
16. **Retrieving:** calling up information from long-term memory into short-term memory.
17. **Representing:** creating a graphic or pictographic mental or visual representation of information.
18. **Summarizing:** combining information parsimoniously into a cohesive statement.
19. **Transposing:** translating information from one code to another.
20. **Valuing:** identifying and articulating personal values relative to information.
21. **Verifying:** confirming or disconfirming the truth of something.
22. **Visual matching:** linking a picture or symbol with a linguistic label.

Fig. 1. General Cognitive Processes

"If students know the concepts . . . contained in the items, they do well on the tests."

different pieces of information. Figure 2 lists the percentages of occurrence of the nine general cognitive operations for each battery.

These findings seem to support the assertions of Costa (1985) and Beyer (1985) that classroom practices, in general, and standardized tests, in particular, do not support many of the important information processing and utilization skills that are necessary for the "information age." For example, Gardner (1983) lists different types of intelligence that are not reinforced in the traditional classroom setting. Similarly, many of the elements of Sternberg's (1985) triarchic model of intelligence are not formally addressed within traditional schooling.

Given that the thinking skills listed in Figure 1 are based on the idea that formalized schooling, in both instruc-

"Test items can be constructed to emphasize underlying cognitive operations rather than declarative content."

tion and assessment practices, does not cover many of the important general cognitive operations, it is no wonder that they are not well represented on standardized test batteries. In other words, the first finding of the study is not surprising.

The second finding of the study, however, is more difficult to explain. Specifically, we found that the general cognitive operations identified on the two batteries accounted for less than 3 percent of the variance in student achievement on the Stanford and 4 percent of the variance on the CTBS. These percentages were calculated using multiple regression analysis, in which one develops an equation to predict the difficulty of each item based on the presence or absence of the nine general cognitive operations. With this measure, if the nine general operations have a strong relationship with item difficulty, then items that contain these operations will be more difficult than items that do not. The strength of relationship between the general cognitive operations and item difficulty is measured in terms of "predictable variance."

In this case, the predictable variance was very small—less than 3 and 4

percent, respectively, for the Stanford and CTBS. In other words, less than 98 percent of the variance in student performance on the Stanford and 97 percent of the variance in student performance on the CTBS are a function of something other than the ability to use the general cognitive operations found in the tests. Such a finding is not consistent with information processing models (e.g., Anderson 1982) that stress the importance of general cognitive operations in processing different types of information. There are several possible explanations.

Three Interpretations

One explanation is that the nine general thinking skills are important and perhaps necessary to complete items on standardized tests but that students have already mastered these skills by the time they take such tests. That is, by the time students begin taking standardized tests they are already adept at comparing and contrasting, summarizing, and the like. Consequently, students use the nine general cognitive operations when answering test items on the Stanford and CTBS; but since virtually all students are adept at them, their presence or absence in the mental processes of answering a test item

has little to do with whether students find the item difficult. This interpretation is supported by the research of Anderson (1983), Fitts (1964), and LaBerge and Samuels (1974), who state that cognitive skills can be learned to a point where they can be performed with little or no conscious effort.

Another interpretation of the second research finding is that the general cognitive operations do not exist independent of the content on which they are used. That is, it is impossible to discuss comparing and contrasting in isolation from the information being compared and contrasted. From this perspective, generic lists of "thinking skills," such as those developed by Costa (1985), Marzano et al. (1987), and Ennis (1985) have little practical value because the skills have no meaning by themselves. This perspective is supported by theorists such as Resnick (1983 and in press) and Glaser (1984).

A third interpretation is that general cognitive operations found in the two achievement batteries are low-level examples of those skills. That is, the nine general cognitive operations can be performed at levels ranging from very simple to quite complex. This interpretation is supported by the work of developmental psychologists such as Fischer (1980) and Case (1985), who assert that most cognitive operations can be performed at differing levels of complexity.

We believe the third interpretation is the most justifiable because the difficulties of some specially constructed test items are a function of the underlying cognitive operations. For example, using a Piagetian model to design science tests, Adey and Harlen (1986) found that the developmental level of a test item was a highly reliable predictor of the item's difficulty. Similarly, O'Brien (1986) found that specially constructed mathematics test items could be reliably calibrated based on the difficulty of the underlying cognitive operations. These efforts and others indicate that test items can be constructed to emphasize underlying cognitive operations rather than declarative content.

Item A

Where would you expect to find a suburb?

1. near a city
2. in a desert
3. on the top of a mountain
4. in a forest

Item B

Which of the following would be the most difficult to learn in two months?

1. playing a piano
2. combing your hair
3. tying your shoes
4. riding a bicycle

Fig. 3. Sample Items

"In the absence of tests that assess underlying cognitive operations, teachers must rely on more informal techniques for analyzing student cognitions."

	Stanford	CTBS
Retrieving	100% N=3,775	100% N=3,167
Reference	17% N=635	27% N=839
Comparing/ Contrasting	100% N=3,775	100% N=3,167
Summarizing	2% N=92	4% N=115
Inferring	6% N=215	7% N=224
Ordering	6% N=241	5% N=150
Visual matching	11% N=420	6% N=204
Transposing	4% N=167	6% N=200
Representing	7% N=266	3% N=81

Fig. 2. Percentage of Occurrence of Nine General Cognitive Operations

“Districts that wish to perform well on standardized tests would do better to teach the facts contained in tests rather than teach generic thinking skills.”

Implications

The implications of this interpretation are profound. First, it implies that standardized tests in their present form are primarily measures of factual or declarative information. To illustrate, consider the two items in Figure 3 (close approximations of two items found in the batteries).

Item *A* is fairly straightforward from the perspective of underlying cognitive operations. To answer it, students must simply retrieve information from their long-term memories about the concept *suburb*—specifically, information about the location of suburbs. They must then compare and contrast that information with the alternatives given.

Item *B* is a little more complex from the perspective of underlying cognitive operations. To answer it, students must retrieve from their long-term memories information about the processes of playing a piano, combing

your hair, tying your shoes, and riding a bicycle. Students must then infer how the difficulty of learning each of the retrieved processes would change under the time constraint. Finally, students must order the retrieved processes based on the newly inferred difficulty level and then compare and contrast their selections with the alternatives.

Even though item *A* requires fewer general cognitive operations and is less complex in terms of manipulating information, it was far more difficult for students. Only 45 percent of the students who took item *A* on the norming sample answered it correctly, while 85 percent of students who took item *B* answered it correctly on the norming sample. Our interpretation is that students had difficulty with item *A* simply because they did not know the concept *suburb*.

This low-level factual or declarative information has little to do with thinking ability: either you have learned the concept *suburb* incidentally or through direct instruction, or you have not. Perhaps the 98 percent and 97 percent of the variance not accounted for in the Stanford and CTBS, respectively, are simply a function of knowing the factual information in those tests. If students know the concepts, principles, and schemata contained in the items, they do well on the tests. If they do not know that information, they do not do well. If this is the case, school districts that wish to perform well on standardized tests would do better to teach the facts contained in standardized tests rather than teach generic thinking skills.

Another implication of our interpretation is that standardized tests should be restructured to emphasize more complex process knowledge in lieu of factual or declarative knowledge. There have already been efforts to do this. For example, considerable progress has been made in developing test items that assess underlying cognitive operations in reading (Valencia and Pearson 1987). Instead of the single-answer, multiple-choice format of current standardized tests, multiple answers are permitted; factual knowl-

edge is accounted for and parceled out of the assessment of general cognitive operations; and probing of student answers is a major component of the assessment process. That is, paper and pencil testing is supplemented by interactions with individual students.

Finally, our interpretation implies that in the absence of standardized tests that do assess underlying cognitive operations, teachers must rely on more informal techniques for analyzing student cognitions. Here, too, there is precedent within education. For example, in assessing reading competence, Goodman (1978) has recommended the use of “kid-watching,” which involves observing, interacting, documenting, and interpreting. Teachers observe students reading in natural settings; they interact with students to gather clues to their use of specific cognitive operations; they record these interactions to help them judge student strengths and weaknesses in cognitive operations. Assessment, therefore, occurs not only at testing time; it is an ongoing part of the teaching/learning process.

“Standardized tests were originally just one of many pieces of data used to assess student competence.”

"The need is clear for . . . alternatives that examine . . . thinking skills important for the information age."

Broader View of Assessment

In a similar vein, Campione and Brown (no date) have developed a technique called "dynamic assessment" that can be used on any set of cognitive operations. Building on Vygotsky's concept of "zone of proximal development," their technique presents students with gradually more explicit cues for performing a task. The initial hints are very general, but become progressively more specific and concrete, with the last "hint" actually providing a detailed blueprint of the cognitive operations required to generate a correct answer. The metric of assessment is how many hints must be provided until students can perform the cognitive operations necessary to complete the task.

The use of more qualitative assessment techniques such as those mentioned above would require a dramatic shift from the narrow view of assessment as a matter of administering paper and pencil tests, to a broader view of assessment as an array of both qualitative and quantitative techniques. According to Haney (1984) such a shift would not at all violate the original intent of standardized tests. Tracing the history and development of standardized tests in this

country, Haney paints a picture of gradual reliance on standardized tests as the major criterion measure of performance and ability. According to Haney, standardized tests were originally just one of many pieces of data used to assess student competence.

The need is clear for restructuring and developing alternatives to standardized tests—alternatives that examine the vast array of thinking skills important for the information age. The present system, with its emphasis on factual declarative content, is out of balance. If this imbalance continues, many students will enter the information age without formal training in necessary skills. □

References

Adey, P., and W. Harlen. "A Piagetian Analysis of Process Skill Test Items." *Journal of Research in Science Teaching* 23 (1986): 707-726.

Anderson, J. "Acquisition of Cognitive Skills." *Psychological Review* 89 (1982): 369-406.

Anderson, J. *The Architecture of Cognition*. Cambridge, Mass.: Harvard University Press, 1983.

Beyer, B. "Practical Strategies for the Direct Teaching of Thinking" In *Developing Minds: A Resource Book for Teaching Thinking*, edited by A. Costa. Alexandria, Va.: Association for Supervision and Curriculum Development, 1985.

Campione, J. C., and A. L. Brown. *Dynamic Assessment: One Approach and Some Initial Data*. Champaign, Ill.: University of Illinois, n.d.

Case, R. *Intellectual Development: Birth to Adulthood*. New York: Academic Press, 1985.

Costa, A. "Teacher Behaviors that Enable Student Thinking." In *Developing Minds: A Resource Book for Teaching Thinking*, edited by A. Costa. Alexandria, Va.: Association for Supervision and Curriculum Development, 1985.

Ennis, R. H. "Goals for a Critical Thinking Curriculum." In *Developing Minds: A Resource Book for Teaching Thinking*, edited by A. Costa. Alexandria, Va.: Association for Supervision and Curriculum Development, 1985.

Fischer, K. W. "A Theory of Cognitive Development: The Control and Construction of Hierarchies of Skills." *Psychological Review* 87 (1980): 477-531.

Fitts, P. M. "Perceptual-Motor Skill Learning." In *Categories of Human Learning*, edited by A. W. Melton. New York: Academic Press, 1964.

Gardner, H. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books, 1983.

Glaser, R. "Education and Thinking: The Role of Knowledge." *American Psychologist* 39 (1984): 93-104.

Goodman, Y. "Kid Watching: An Alternative to Testing." *National Elementary School Principal* 57 (1978): 41-45.

Haney, W. "Testing Reasoning and Reasoning About Testing." *Review of Educational Research* 54 (1984): 597-654.

LaBerge, D., and S. J. Samuels. "Toward a Theory of Automatic Information Processing in Reading Comprehension." *Cognitive Psychology* 6 (1974): 293-323.

Marzano, R., and D. Jesse. "A Study of General Cognitive Operations on Two Achievement Batteries and their Relationship to Item Difficulty." Unpublished paper, Mid-continent Regional Educational Laboratory, Aurora, Colo., 1987.

Marzano, R., R. Brandt, C. Hughes, B. F. Jones, S. Rankin, B. Presseisen, and C. Suhor. *Dimensions of Thinking*. Alexandria, Va.: Association for Supervision and Curriculum Development, 1987.

O'Brien, M. L. "Calibrating Item Difficulty as the Basis of Prescriptive Test Theory." *Studies in Educational Evaluation* 12, 1 (1986): 5-29, 59-69.

Resnick, L. B. "Toward a Cognitive Theory of Instruction." In *Learning and Motivation in the Classroom*, edited by S. Paris, G. Olson, and H. Stevenson. Hillsdale, N.J.: Erlbaum, 1983.

Resnick, L. B. *Education and Learning*. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh, in press.

Sternberg, R. J. *Beyond IQ: A Triarchic Theory of Human Intelligence*. New York: Cambridge University Press, 1985.

Valencia, S., and P. D. Pearson. "New Strategies for Reading Comprehension Assessment." Unpublished paper, Center for the Study of Reading, Champaign, Ill., 1987.

Robert J. Marzano is Director of Research, Mid-continent Regional Educational Laboratory, Aurora, CO 80014. **Arthur L. Costa** is Professor, California State University, School of Education, 6000 J St., Sacramento, CA 95819-2694.

Copyright © 1988 by the Association for Supervision and Curriculum Development. All rights reserved.