Teaching and Testing to Develop Fluid Abilities

DAVID F. LOHMAN

Fluid abilities are important both as aptitudes for success in formal schooling and as outcomes of formal schooling. However, the aptitude function has been overemphasized, and the outcome function ignored altogether, primarily because fluid abilities are often mistakenly thought to be innate. Two methods for developing and assessing these abilities are discussed. In the first method, students are asked to solve increasingly unfamiliar problems in a domain. This usually requires the adaptation of existing problem-solving strategies or the assembly of new strategies to solve increasingly ill-structured problems. In the second method, students are required to organize knowledge in new ways or to view it from different perspectives. In both cases, assessment requires that students' personal perspectives be elicited. This raises several difficult philosophical and psychometric problems. These problems and the implications for education are discussed.

Educational Researcher, Vol. 22, No. 7, pp. 12–23

Cronbach (1975) once observed that strange ironies abound in the history of mental testing. For example, Terman's (1925) longitudinal study, one of the most influential investigations of academically gifted students, was conducted with his adaptation of Binet's test, a test that was originally designed to identify academically retarded youth. Wechsler's (1939) test, originally developed to provide a better measure of adult intelligence, became the most popular children's test and was even adapted for use with preschool children. One of the greater ironies, though, is the way in which advances in intelligence testing may have inadvertently thwarted educational practices that encourage the development of abilities measured by these tests, particularly those we now call fluid abilities. This article discusses how this has happened and suggests instructional and measurement strategies that encourage the development of fluid abilities.

Understanding Fluid Abilities

Age-Normed Scores

Much of the confusion in our understanding of abilities stems from the use of age-normed scores. An example may help. Suppose a child is administered an IQ test every year for 4 years. The first test is administered when the child is 6 and the last test when she is 9. We observe IQ scores of 150, 143, 137, and 133. Would we say her ability is increasing or decreasing? Clearly, it seems to be decreasing. Suppose, however, that we look at the mental age (MA) scores that correspond to these IQ scores. Note that MA scores are much like a total number correct on a test, or grade-equivalent scores on an achievement test. What we then see are MA scores of approximately 9, 10, 11, and 12. Is ability increasing or decreasing? Clearly, the child is getting better scores on the test each year. Indeed, even her rate of development is average: Each year she gains exactly 1 year on the MA scale. The declining IQ score tells us only that her rate of growth lags behind others of the same rank within her age group.

These are ancient facts in the history of mental testing. Investigators have often found negative correlations between age and IQ for children who live in areas where schooling is inadequate or nonexistent. For example, in his study of children of gypsies and boat canal pilots in London, Gordon (reported in Freeman, 1934) found a correlation of $r = -.75$ between IQ and age. Young children showed MA scores only slightly below that of the average London schoolchild. By the time these children reached 16 years of age, however, average MA scores had increased to 9.6 but average IQ had plummeted to 60. Sherman and Key (1932) and Sharp, Cole, and Lave (1979) provide other examples. The point is not that some environments are bad and others are good. Rather, the point is that in order to keep one's place on age-normed scale, one must get better each year at the same rate as others with the same initial score. Just getting better will usually not do.

It is sometimes difficult to see this for broad abilities such as those assessed by intelligence tests. Because of this, many erroneously think that the year-to-year gains observed on such tests are the product of normal maturation in a normal environment. Hunt (1961) long ago challenged this notion, but it persists, perhaps because it conforms to our naive theories about what abilities are and how they develop. Abilities do not mature of their own accord. Both the type and amount of schooling are predictors of gains or losses in IQ scores (see Hunt, 1961; Snow, 1982, for reviews). Thus, as Snow (1982) puts it: "Educational psychology now recognizes intelligence as education's most important product, as well as its most important raw material" (p. 496).

Because all abilities are developed through experience and exercise, I find it much more informative to look at growth curves for raw scores than to covary age out of the picture. This does not mean that abilities develop at the same rate or approach the same limits in all individuals. Indeed, biological factors are important, although the limits set by biology are soft, vary over time, and are unknowable in ad-

DAVID F. LOHMAN *is professor of educational psychology and chair, Division of Psychological and Quantitative Foundations, 366 Lindquist Center, University of Iowa, Iowa City, IA 52242.*

vance. Further, the interactions between heredity and environment seem to be more complex than we once imagined. For example, Rice, Fulker, DeFries, and Plomin (1988) found no significant correlation between a child's IQ at age 7 and aspects of the home environment at ages 3 and 4. However, they did find significant correlations between general intelligence at age 7 and aspects of the home environment at ages 1 and 2. The timing of environmental events may thus be more important than their presence or absence (Ceci, 1990).

In summary, beliefs about the origin of abilities conflict with beliefs about the need for educational resources to develop those abilities. For example, the only real justification for something called gifted education is the recognition that abilities do not mature of their own accord and that the promise of great achievement can easily be squandered if needed educational resources are not provided at the appropriate time. The *possibility* of becoming a great chess player, writer, musician, mathematician, or scientist may be in significant measure a gift from one's ancestors. But the *attainment* of excellence comes only after much training and practice.

But how can we identify those who show promise for great achievement later? Historically, intelligence and other aptitude tests have served this purpose. Advocates such as Burt (1955) and Terman (1916) generally believed that they were assessing something like the innate potential or capacity of the learner with their IQ tests. Some critics thought that existing tests did not do this well, but often assumed that a new test could some day be developed that would provide an unbiased estimate of learning potential. Other critics seemed to accept that IQ tests measured potential or capacity, but argued that it was really only a very narrow sort of potential. What was needed was a more democratic approach to assessment in which other potentials could also be identified (e.g., Thurstone, 1938; Gardner, 1983). In my opinion, all of these views are misleading. All assume that we can identify innate potential—whether in the singular or in the plural. Intelligence tests, like achievement tests, measure developed abilities, not innate capacity or potential.[1]

The Nature of Human Abilities

Abilities can be understood as transferable knowledge and cognitive skills. Broad abilities are defined either by homogeneous samples of knowledge and skill that transfer widely (e.g., metacognitive knowledge, working memory) or by heterogeneous samples of knowledge and skill that transfer more narrowly (e.g., general academic achievement). Specific abilities are defined by homogeneous samples of knowledge and skill that transfer narrowly. Abilities may thus be ordered in a rough hierarchy. Near the apex of the hierarchy are broad abilities that predict performance in many situations. Thousands of specific abilities form the base.

Correlational studies of human abilities show that intelligence tests (particularly the so-called performance variety) often measure something Cattell (1963) and others call fluid ability (Gf). General academic achievement tests, on the other hand, usually measure something Cattell (1963) calls crystallized abilities (Gc). I believe that we have learned something important about both of these constructs that could improve what we assess, and thus what students and teachers see as the goals of instruction. In a nutshell, here is my first argument: Fluid abilities are among the most important aptitudes for learning as well as one of the more important outcomes of education. However, existing tests of fluid abilities have emphasized the aptitude function and ignored the outcome function. Such tests have little instructional utility. This would not matter except for the fact that untested outcomes tend to go untaught. Children are thus less able to transfer their learnings to new situations than if we had instructionally useful ways to assess fluid abilities. This is shown in Figure 1.

The horizontal line here represents the amount of transfer required by problems or their novelty to the test taker. Tests at the far left consist of problems that are duplicates of those taught. As one moves to the right on this scale, problems become increasingly novel and thus require increasing transfer. This continuum or something like it has been proposed by many investigators. I thought that I had discovered it in the mid-1970's, but then I found it in Cronbach's (1970) text, then Anastasi's (1937) text, then in Stern's (1914) monograph, then in Raaheim's (1984) work, then in a paper by Elshout (1983), and finally in Sternberg's (1985) triarchic theory. Alliteratively, then, one can find it from Stern to Sternberg.

Putting "intelligence" or fluid abilities and achievements or "crystallized abilities" on the same line implies that these are best seen as two aspects of the same thing rather than as qualitatively different things (Snow, 1980). In fact, Snow (1980) speaks of fluidization and crystallization processes to describe these different aspects of ability development. He summarizes:

> [Crystallized ability] may represent prior assemblies of performance processes retrieved as a system and applied anew in instructional or other . . . situations not unlike those experienced in the past, while [fluid ability] may represent new assemblies of performance processes needed for more extreme adaptations to novel situations. The distinction [then] is between *long term* assembly for transfer to *familiar* new situations vs. *short term* assembly for transfer to *unfamiliar* new situations. Both functions develop through exercise, and perhaps both can be understood as variations on a central production system development. (p. 360)

Figure 2 shows how this crystallized-to-fluid continuum can be applied to the task of adding numbers. Suppose students learn to do this on paper rather than in their heads. Also, suppose that during learning, numbers are always presented in a column format. Presenting similar problems in a row requires some transfer. Embedding addition prob-

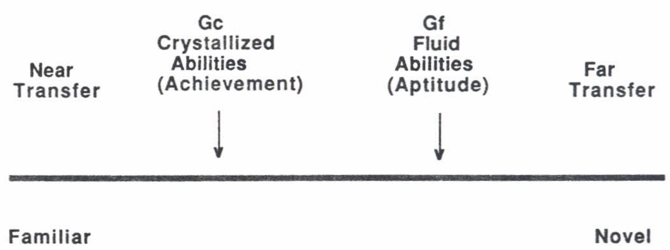

FIGURE 1. *A continuum of task novelty or transfer applied to the construct intelligence.*

lems in a story format requires more transfer (and additional skills). Embedding them in matrix problems such as the Progressive Matrices test of Raven (1962) requires even more transfer. Insight problems of the same sort may require the most transfer.

This continuum can apply to other ability constructs as well. For example, consider musical abilities. One can teach students to play a musical instrument in a highly structured way; such students often show little ability to be innovative or to compose new pieces themselves. We often see the opposite emphasis in art, where students are given paints and told to be creative, but given no structure or direction. It also applies at the most general level to summary measures of school competence. This is shown in Figure 3. Here typical sorts of educational assessments are placed along the line. Although daily quizzes (on the far left) may require considerable transfer, mostly they require much less. This is in part an avoidable consequence of the way such tests tend to be constructed and in part an unavoidable consequence of the fact that it takes time to consolidate and apply newly learned knowledge and skills. I will return to this point later. Next come the more general tests in which students must transfer their learnings over time and to increasingly unfamiliar problem types. Note again that I have placed tests of fluid abilities (or, as some would call it, intelligence) on this line as well. These tests provide one way to estimate how well students can transfer their learnings to unfamiliar situations. However, scores on such tests are not instructionally transparent, and so this is not a very good way to assess how far, in general, students can transfer their learnings.

Several theorists have defined intelligence or scholastic aptitude as the ability to profit from incomplete instruction—that is, the ability to transfer old learning to new situations. Older and more able students are more likely to show transfer than are younger and less able students (see, e.g., Campione, Brown, & Ferrara, 1982). Thus, both "intelligence" and "transfer" are descriptors of the same phenomenon. Unless we view intelligence as fixed or transfer as unattainable, then the road to achieving transfer is also the road to developing intelligence.

However, like intelligence, "transfer" is a multidimensional concept that is only imperfectly captured by a single dimension. For example, Gagne (1970) distinguished between vertical and lateral transfer. Vertical transfer is said to occur when the acquisition of a subskill contributes directly to the acquisition of a superordinate skill of which it is a subordinate. Lateral transfer refers to "how broadly the individual can generalize what he has learned to a new situation" (Gagne, 1970, p. 336). Historically, intelligence has been more associated with vertical transfer and creativity

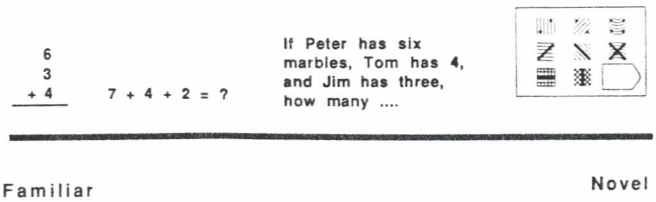

FIGURE 2. *A transfer continuum for addition problems.*

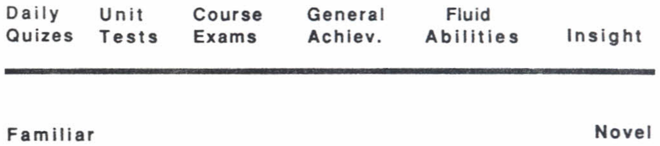

FIGURE 3. *A transfer continuum applied to typical educational assessments.*

with lateral transfer. I would argue, however, that vertical transfer is better associated with average mental development (something like Binet's original concept of mental level or Piaget's adaptation of this idea) whereas lateral transfer is concerned with the relative crystallization or fluidization of these abilities. Thus, one can chart the development of any ability—broad or narrow—along two major dimensions: (a) the degree of vertical transfer achieved, which will be highly related to the abstractness of the concepts and skills developed, and (b) the range of lateral transfer achieved, that is, the breadth of the domain in which such knowledge and skills can be applied. Those concerned with general mental development have naturally emphasized the vertical dimension of transfer. I am here more concerned with the fluidization of abilities and thus focus on lateral transfer.

The Fluid-Crystallized Distinction

The distinction between fluid intelligence and crystallized intelligence lies at the heart of the argument that fluid abilities are as much the product of education and experience as are crystallized abilities. It is interesting to note how these concepts have evolved. When Cattell (1943) first proposed his theory, he emphasized the equality of fluid and crystallized abilities. However, in later versions of the theory (Cattell, 1963, 1971), fluid ability was interpreted as something like the true, innate intelligence of the individual that, when invested in experience, produced a particular mixture of crystallized abilities. The investment theory of aptitude has considerable intuitive appeal. However, the problem is that there is no way to measure innate potential or capacity. Fluid ability can be defined in this way, but it cannot be measured.[2]

Indeed, most educators and many psychologists think intelligence tests measure—or ought to measure—something like the innate capacity or potential of the learner. This has always been a popular belief among both professionals and laymen. It is a personal theory that is staunchly held and, like other personal theories, is not easily altered by disconfirming evidence (e.g., Champagne, Klopfer, & Anderson, 1980). Those who hold such theories either openly advocate the search for better measures of what they consider to be the true, physiological intelligence of individuals (as in the work of Eysenck, 1982, or Jensen, 1982) or, conversely, espouse an environmentalist explanation for intelligence and reject existing intelligence tests as biased because they show differences between ethnic groups and social classes. Paradoxically, these same critics of existing tests often assume that an unbiased measure of intelligence could be found that would not be influenced by differences in education, social class, motivation, or the like. Humphreys (1986) aptly calls such people "closet hereditarians." Like politics, personal theories make strange bedfellows.

Much evidence challenges the typical naive theory of intelligence as an innate capacity (see Angoff, 1988, for one summary). For example, if intelligence tests measured something more innate or biologically based than achievement tests, then one would expect heritabilities to be higher for intelligence test scores than for achievement test scores. Although some investigators have found higher heritability coefficients for intelligence tests (e.g., Cattell, 1982; Thompson, Detterman, & Plomin, 1991), others have found that achievement tests show equal or even higher heritabilities (Horn, 1985; Humphreys, 1981; Scarr & Carter-Saltzman, 1982). Humphreys (in Davis & Flaherty, 1976, p. 211) once noted that in the Project Talent data, heritability coefficients were higher for a test of Bible knowledge than for a test of nonverbal reasoning or fluid ability.[3]

A second line of evidence that has supported the view that fluid abilities reflect the biological integrity of the brain have been studies that show general fluid and crystallized abilities are affected by different factors over the life span, with fluid abilities showing the greater decline. For example, Horn, Donaldson, and Engstrom (1981) related declines in fluid ability with aging to declines in attention control and working memory. Similarly, Salthouse (1985) argues that declines in fluid ability reflect declines in working memory resources available.

General fluid abilities at all ages may be constrained by the ability to transform information in working memory (in fact, Kyllonen & Christal, 1990, argue that the constructs are synonymous). Declines in working memory with age may indeed be caused by cumulative damage to the central nervous system. However, these declines may also in part reflect unfamiliarity with schoollike tasks or limited practice in exercising such abilities that results either from lack of opportunities or from an increasing reliance on stored knowledge and solution strategies. Furthermore, if working memory is not so much a place as a state of activation (Anderson, 1983), and if activation spreads automatically to related nodes in memory, then declines in the ability to focus attention may be a necessary consequence of (a) the acquisition of an increasingly diverse knowledge base and (b) the repeated activation and use of that same knowledge base. This implies that such declines could be attenuated by systematic attempts to move outside the bounds of the familiar, to learn new concepts and skills rather than continue to entrench the old (cf. Sternberg & Gastel, 1989).

The important point, however, is that the constructs of general fluid and crystallized intelligence need to be distinguished from the process whereby particular abilities are made more fluid or more crystallized. Fluidization of any ability may indeed require greater working memory resources than crystallization of that ability, since fluidization requires that learners revise existing problem-solving strategies, assemble new ones, search for new analogies or new perspectives, and the like. Furthermore, one can estimate general fluid ability either by averaging across several puzzlelike tasks that require subjects to transform familiar stimuli in working memory (as in Kyllonen & Christal, 1990), or by averaging across a much wider sample of schoollike tasks that require such thinking in different content domains. Those interested in estimating aptitudes may prefer the former route, whereas those interested in developing such abilities may find the latter route more profitable.

Fluid Abilities as Outcomes of Education and Experience

Although I would agree that existing tests of intelligence are not terribly useful for educators and can even be harmful (primarily because of the mistaken belief that they measure potential or capacity), I would argue that they measure one of the most important *outcomes* or products of education. What we need are better, more instructionally useful ways to estimate the ability to solve unfamiliar, ill-structured problems. In other words, we need measures that show high correlations with old tests of fluid abilities but that are instructionally useful. More importantly, we need these measures not only at the level of general outcomes of education but also at the level of more specific abilities that are the object of instruction.

In the past 20 years, there has been much research that has sought to understand the nature of human intelligence as assessed by intelligence tests. Sometimes intelligence was treated as an end in itself, as in Sternberg's (1985) work. But others, notably Snow (1978) and Glaser (1972), claimed that the ultimate objective of their research on "intelligence" or scholastic aptitude was to understand better why some students learned well in school while others failed to learn. In other words, they were attempting to go from right to left in Figure 3. Although much was learned from these efforts, I believe that we will make better progress going the other way, that is, by devising ways to estimate how far students can transfer their learnings. Although we have important educational objectives all along the transfer continuum, the ability to transfer school learning to unfamiliar situations is perhaps the single most important (measurable) outcome of schooling. Therefore, we must strive mightily to teach and to assess these abilities in all domains of instruction.

Unfortunately, however, measurement problems increase—perhaps exponentially—as we move from left to right on this scale. Tests that sample no more than those facts and skills explicitly taught are relatively easy to defend. Tests that require transfer are more difficult to defend because problem novelty varies across individuals and over time. In other words, as we move to the right, measurement becomes increasingly probabilistic and the inferences based on tests scores become increasingly tenuous. Some argue that defensible tests of insight (creativity?) on the far right of the continuum are nonexistent.

Again, this would not be a problem except for the fact that unmeasured outcomes often go untaught (Frederiksen, 1984). If fluid abilities are indeed in some measure the product of experience, then educational programs that promote or inhibit the solving of unfamiliar problems should have a measurable impact on the development of these abilities. In fact, there is some evidence for this in Kennedy's (1978) evaluation of project Follow Through (see Snow, 1982). Highly structured programs such as the Direct Instruction program at the University of Oregon achieved considerable success in developing crystallized abilities of students. These programs required students to engage in much stylized drill and practice. However, gains on a fluid ability test for these children were among the lowest. Other programs, such as the Responsive Environment Program at Far West Labs in San Francisco, showed the opposite pattern. These programs were relatively unstructured, used learning or concept centers, and focused on general problem solving and school readiness. In other words, program output reflects program input. And those who see tests of fluid abilities as measuring something that is fixed or innate miss evaluating an important outcome of education. Both fluid and crystallized abilities are the products of education and experience.

This is not an isolated finding. There is a substantial literature on the effects of schooling and occupation on changes in abilities and ability patterns. Hunt (1961) reviewed much of the early work. But there are many other studies. For example, Balke-Aurell (1982) examined ability changes as a function of education and occupation for two 10% random samples of Swedish males. The findings were quite clear and remarkably consistent in both samples: (a) A verbally demanding educational experience produced larger gains in verbal ability; (b) a technical or vocational educational experience produced larger gains in spatial ability; (c) verbally and spatially demanding occupational experience had similar effects on the pattern of verbal versus spatial ability development; and (d) more education produced greater gains in general intelligence.

Although the finding that intelligence improves with schooling is certainly important, my main concern here is with the prior findings that the type of education received can have an important impact on the nature of the ability profile as well. I would also argue that the educational system that emphasizes skill and fact learning produces a different type of student than the system that encourages exploration and discovery. Indeed, this may be why changes in intelligence test scores over the past generation have generally been larger on tests of fluid abilities than on tests of crystallized abilities. Such a result is difficult to explain if fluid abilities are less influenced by environment than crystallized abilities. Flynn (1987) explains it by assuming that "IQ tests do not measure intelligence but rather a correlate with a weak causal link to intelligence" (p. 171). Others are less willing to dismiss IQ tests. For example, Lynn (1990) argues that increases in intelligence reflect improvements in nutrition rather than improvements in education. Brody (1992) agrees that nutrition may be partially responsible, but notes that such an argument really does not explain why educational interventions (e.g., Stankov, 1986) may produce larger changes in fluid ability than in crystallized ability. Such findings are not so difficult to explain if we view fluidization as a process. Educational and social changes during the past generation were generally not of the sort that, when compared with educational and social conditions of the previous generation, would foster greater crystallization of abilities developed through formal schooling, but rather might be expected to encourage their fluidization. Indeed, one might use Flynn's (1987) data to argue that the educational reforms implemented during this period worked better than many thought, especially those who focused on declines in achievement test scores.

Understanding Over- and Underachievement

Understanding that fluid abilities are developed, not innate, can suggest how we might assist all students in developing their abilities. In other words, one can plot programs or classes or schools in a space defined by fluid and crystallized abilities (e.g., Snow, 1982). One can also plot individuals in such a space, as in Figure 4. Traditionally,

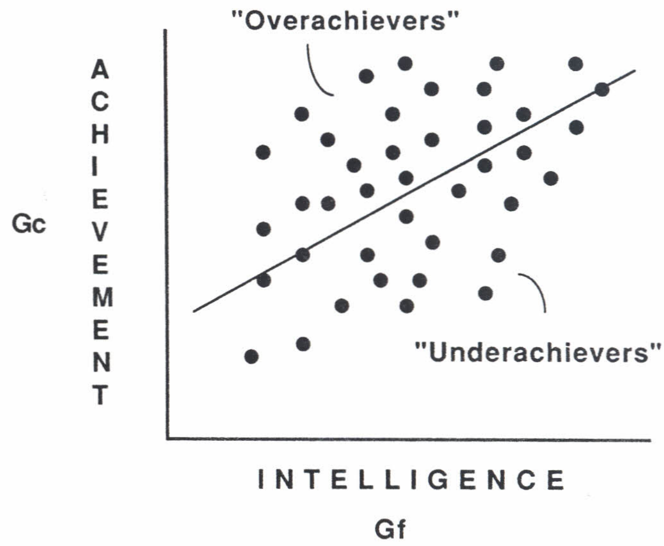

FIGURE 4. *Plot of intelligence (or Gf) versus achievement (or Gc). "Overachievers" are above the regression line; "underachievers" are below the regression line.*

such plots were used only to identify so-called underachievers, that is, those students who needed more subject-matter learning. However, the concepts of over- and underachievement presume that fluid ability (or intelligence) drives, and thus sets limits on, crystallized ability (or school achievement) (see Thorndike, 1963). This conforms nicely to intuitive theories of intelligence, although it conflicts with the data. If we interpret the scores of students who fall below the regression line as evidence that they are not achieving at a level commensurate with their potential, then are those students above the regression line achieving at a level higher than their potential? Is it possible to do better than one can? Similarly, we know what to prescribe when the student falls below the regression line: more study or better education. But what, then, do we tell the other half of the students who fall above the regression line? Do we tell them to study less? These so-called overachievers also need to learn something: They need to learn how to apply what they have learned to unfamiliar situations. Perhaps they need guided practice in confronting and coping with novelty. Or perhaps they need to be taught in ways that encourage the development or more flexible cognitive units, units that can be assembled and reassembled to solve increasingly ill-structured problems (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987). Thus, instead of speaking of overachievers and underachievers, as is still frequently done in the educational literature, it is much better to speak of students who are relatively higher on fluid applications than crystallized applications or vice versa. This distinction, then, can be applied not only at the level of general abilities but also at the level of narrower abilities, such as mathematical achivement, or even specific abilities, such as knowledge of algebra.

The traditional explanation for overachievement or underachievement has focused on the role of motivation: The overachiever is too motivated whereas the underachiever is not sufficiently motivated. Although this is not a complete explanation, motivation or personality variables can play an important role in ability development and may also involved in the development of individual and group differences in some abilities. Consider two students. The first always does her homework, and as she progresses through school, routinely adapts her thinking and behavior to conform to whatever is expected by the instructor. The second student often does not do his homework, thus often is not prepared for class, and when called upon must quickly invent an answer that fits the situation, usually by grasping at prior knowledge that seems related to the question posed.

The hypothesis, then, is that the conforming student will more likely show higher development in crystallized applications, whereas the unprepared student will more likely show higher development in fluid applications. In other words, "underachievers" may score higher on tests of fluid abilities than on tests of crystallized abilities in part because they have had more opportunities to exercise and develop these abilities, abilities that the well-prepared students did not need to exercise as often. Further, sex differences in conformity to school norms may in part account for the fact that boys often score higher on achievement tests that present moderately novel problems whereas girls receive higher grades (Halpern, 1992).

However, sex differences are not the point here. Rather, the point is that different educational experiences result in different patterns and levels of ability development. One's educational experiences vary because of external factors, such as the nature of the curriculum, the instructional methods, or the evaluation measures used, or because of internal factors, such as one's conformity, assertiveness, and the like. But whatever the reason, "overachieving" students—that is, students who score much higher on tests of crystallized knowledge than on tests of fluid applications of that knowledge—also need to learn something. A closer look at transfer and its relationship to fluid abilities suggests what to teach such students, and how.

Toward Instructionally Useful Measures of Fluid Abilities

How can one assess fluid abilities in instructionally useful ways? This can be done in at least two ways: (a) by estimating a student's ability to adapt existing problem-solving procedures or assemble new ones to solve increasingly novel problems and (b) by estimating the flexibility of the organizational schemes a student can impose on factual knowledge in a domain. The first avenue emphasizes the skill or procedural aspect of knowledge; the second avenue emphasizes the factual or declarative aspect of knowledge.

Fluidization Through Novelty

The crystallized-to-fluid continuum presented in Figures 1 to 3 exemplifies the first route. Here the goal is to start with the known or familiar and push toward the unknown. This is much easier said than done, however, because the scale along which tasks are arrayed becomes fuzzier with each step taken in the direction of increasing novelty.

This was demonstrated quite nicely in two recent studies of mathematical problem solving done by Janey (1991). Janey first sought to determine whether novelty must be defined differently for different individuals. One way to answer this question would be to give students test questions that ostensibly vary in their familiarity and ask each student to rank them according to novelty. If students can perform the task dependably, but give different rank orders, then indeed novelty would appear to be in the eye of the beholder. On the other hand, if correlations among ratings show a large general factor, then task novelty is something that is roughly the same for all students. Janey did this with a sample of mathematics items from the Iowa Tests of Basic Skills (Hieronymous, Hoover, & Lindquist, 1986) and the Cognitive Abilities Test (Thorndike & Hagen, 1986), and then in a second study with algebra items of her own making. Results showed (a) that most students can give dependable ratings of novelty, (b) that most students clearly distinguish between novelty and difficulty, (c) that students do indeed perceive novelty differently (i.e., there is not one factor but several factors in the correlation matrix of their ratings), and (d) that diversity of opinion increases as average novelty increases. Note that the problem here is one of individual differences, not group means. Janey could rank order items according to their average novelty for test takers. However, variance in perceived novelty also increased as average novelty increased. This means that inferences about the meaning of scores for individuals became increasingly suspect as average amount of transfer required by a task increased.

Cognitive science and transfer. We have learned much about transfer in recent years. Using the formalism of production rules, Singley and Anderson (1989) have proposed an updated version of Thorndike and Woodworth's (1901) identical elements theory. Similarity between two tasks is quantified by comparing the number of overlapping and distinctive production rules used to model performance on the separate tasks. An analysis of transfer among text editors supported their model. Although this work is theoretically encouraging, it is unlikely to have a significant impact on practical efforts to assess transfer. Why? First, a detailed task analysis requires an incredible amount of work. Such analyses would not be feasible for a large number of tasks or problems. Second, models vary in the level of abstraction assumed. The mental models used by other theorists (e.g., Butterfield & Nelson, 1989) to account for transfer subsume the production systems of Singley and Anderson (1989). Thus, absolute transfer distances depend on the level of abstraction selected. Third, and most troubling, tasks can be solved in multiple ways. There is no guarantee that the subject represents the task in the same way as the theorist who builds the production system. Indeed, other research suggests that subjects of diverse abilities solve transfer problems in quite different ways (see Snow & Lohman, 1989). Analogical theories of transfer (e.g., Holland, Holyoak, Nisbett, & Thagard, 1987) suffer from the same limitation. In fact, the distinction between training and transfer task is itself possible only if the experimenter has been able to control subject's access to problems of each type. This is not possible outside the laboratory (and is often an unreasonable assumption in the laboratory). In short, modern theories of transfer, like their predecessors, seem unable to untangle the measurement problems caused by the fact that transfer tasks are differentially novel to different problem solvers. Like beauty, novelty is in significant measure in the eye of the beholder.[4]

But if we cannot cleanly manipulate the amount of transfer required by varying stimulus attributes, what are we to do? Behaviorist and cognitivist psychologies differ in many ways, but are united in their assumption that behavior can be systematically related to a common, objective set of stimuli or stimulus attributes. The question of whether my green is your blue is little more than an interesting philosophical query until my behavior differs systematically from yours in response to blue and green stimuli. This is generally not the case when we have both grown up in the same social system, especially when we deal with concepts that are closely tied to our external environment. But when our social experiences differ, even highly familiar concepts may be understood in different ways. It is even more important when we define a construct by the discrepancy between two internal events. The conceptual novelty of a problem is defined by the discrepancy between one's representation of a problem and one's knowledge base assembled from prior experiences.

Taking the problem solver's perspective. The solution to the problem of measuring novelty seems to lie in a rejection of the assumption that test items (or experimental tasks) can be defined by their objective or external characteristics. Instead, if we are to measure transfer, we must somehow elicit the cooperation of the subject and try to see the world from his or her perspective. There is a rich tradition of this sort in the phenomenological psychology in Europe, exemplified in the writings of the philosopher Heidegger (1982) and recently displayed in the work of the Swedish psychologists Marton and Svensson (1979). They summarize the difference between the traditional and phenomenological perspectives in the following way:

> The traditional perspective in research . . . focuses attention on the learner . . . to test hypotheses about how he can be characterized, what he does and how he functions. The learner is the object of our study and we (the researchers) observe him and his behaviour or functioning. We thus observe the learner and describe him as we see him and we observe the learner's world and describe it as we see it. We frequently relate our description of the student to our description of his world and generally do this within an explanatory framework.
>
> There is, however, an alternative perspective we can take: the learner's own. In this perspective the world as experienced by him becomes visible. His experience of the world is a relation between him and his world. Instead of two independent descriptions (of the student on one hand and of his world on the other) and an assumed relationship between the two, we have one description which is of relational character.
>
> There are thus two distinctively different perspectives which we can adopt. . . . One is observational, one is "from-the-outside" . . . and the other is experiential [or] "from-the-inside." The two perspectives are complementary in the sense that neither is derivable from, nor contradicts the other. (p. 472)

My claim, then, is that the measurement of individual differences in transfer requires a shift from the perspective of the test administrator to that of the test taker. And this is why the measurement of transfer has presented such a stumbling block to conventional experimental and psychometric research.

We can gain a glimpse of the test takers' perspectives in a variety of ways, the simplest of which is simply to ask them to rate or sort problems according to perceived novelty. Data that are qualitatively richer but more difficult to classify can be obtained by asking students to explain their understandings of a situation and its implications (or, if a problem, its solutions). Marton and his colleagues have developed a more systematic procedure for doing this, which he calls phenomenography. The aim of the approach is to discover students' conceptions and misconceptions in a knowledge domain and also students' self-perceptions as learners. Typically, students are given open-ended questions that they are asked to solve aloud. Verbal protocols are collected and later analyzed to reconstruct categories and concepts that seem to describe the students' mental experiences. Learning is then viewed as a transition among these alternative conceptual frameworks. The approach has been applied in several knowledge domains, including physics (Marton, 1983), economics (Dahlgren, 1979), and map reading and way finding (Ottosson, 1987).

In summary, although preliminary attempts to bring the perspective of the test taker to bear on the problem of defining what the test measures are encouraging, there are many troublesome psychometric and philosophical problems here. Some of these problems can be avoided if the goal of measurement is to understand what students understand rather than simply to rank order students on a common scale. "What is the student's current conception of the problem?" is a different question from "Who can best solve problems of this type?" The first question presumes a classification; the second, a rank ordering. Psychometric theory from Galton to Rasch has been developed to answer questions of the second sort; only recently have efforts been made to develop a psychometrics capable of addressing the first question (see, e.g., Mislevy, 1993; Lohman & Ippel, 1993). Perhaps these efforts to decouple measurement from individual differences will help. In the end, though, the phenomenological approach may be of greater value to the classroom teacher than to the person charged with the task of producing a summative evaluation, although Janey's (1991) efforts may offer guidance for the traditional test constructor as well.

Fluidization Through Flexible Organization

The second method for encouraging the development of fluid abilities is through instructional activities and testing procedures that encourage students to develop knowledge structures that are richly interconnected (Spiro et al., 1987) but "loosely coupled" (Snow, 1986). As I see it, this divides into two subproblems. The first problem is learning to impose at least one personally meaningful structure on the factual knowledge one learns. The second problem is learning to go beyond this initial scheme, learning then to view concepts from multiple perspectives, thereby developing a rich but loosely interconnected knowledge base.

Imposing an initial organization scheme. There has been increasing concern in recent years with what Norman (1977) calls learning over the long haul. What type of knowledge structures do students assemble over many years of schooling? What are the effects of different foundations on the type of knowledge structure that is eventually assembled? What has happened to students who think they are learning well, who get high grades or marks, but who can later remember very little about what they studied?

Part of the problem, I think, is the way we have come to test recognition instead of recall in routine assessments of student learning.[5] But a larger problem is that tests often not encourage students to identify main ideas when they are learning or to discover ways in which these ideas might be organized in different ways or related to other ideas or beliefs they may have. I am not so much concerned here with nationally constructed tests as with the daily, weekly, and course tests that teachers construct or, increasingly, take from a teacher's manual.

Here is an example of what I mean. Several years ago we celebrated the 200th anniversary of the U.S. Constitution. Children in schools studied special units on the Constitution. I was working with a class of fifth-grade students at the time and was shown their mastery test of multiple choice and true-false questions on this unit. The first thing I noticed was that nowhere on the test was there a question like "What is a constitution?" or "Why do we need a constitution?" Thus, it was possible for children to come away from this test thinking they knew something about the U.S. Constitution without having the foggiest notion of what a constitution is or why a government might need one. Questions on the test ranged from reasonable to absurd. One of the latter was, "Who was the signer of the constitution who had six children?" Answer: Thomas Jefferson. There are many things one might want to know about Thomas Jefferson, but surely the fact that he had six children is not one of the most important. Yet this was the only question about Thomas Jefferson on the test.[6] The problem here is that it tells the child that every idea in the text is equally important. There is no need to separate main ideas from details; all are worth one point. And there is no need to assemble these ideas into a coherent summary or to integrate them with anything else because that is not required.[7]

What is the consequence of this endless stream of disconnected learning? Surveys that sample the factual knowledge of students and adults in our society often show an appalling lack of knowledge about facts that it would seem impossible not to learn after 12 or more years of formal schooling. For example, in a recent Gallup survey of geography knowledge in nine countries, American young adults ranked last. Given a world map, three fourths of the Americans surveyed could not find South Africa or France; nearly one fourth could not find the Soviet Union or the Pacific Ocean. One out of seven could not even find the United States. How does this come about? How could you live in this world and not find the Pacific Ocean, which covers approximately 40% of the globe?

The problem is this: Students learn to study so that they will do well on the test. When the test gives equal weight to the question "Who had six children?" and "What is a constitution?" (if it even has such a question), students do not learn to separate the wheat from the chaff. Nor do they learn how important it is to impose an organizational scheme or two or three on the facts that they must learn. The problem is *not* the learning of facts, but the learning of disconnected facts. Students cannot use what they may know about Piaget's theory in their writing or speaking or thinking unless they can bring it to mind easily. This means that they have organized it.

The claim that there is something amiss with most objective multiple choice and true-false tests is an old one. Yet repeatedly, investigators have found that essay tests—the preferred alternative of the critics—are not much better. Indeed, they are usually less reliable and invariably harder to score. Occasionally investigators report that students study differently for essay tests than for multiple choice tests, but the long-term consequences of learning differently are not explored.

A study by Mittelholtz (1988) offers some suggestions here. Mittelholtz looked at long-term recognition and recall of concepts in an introductory psychology course. One section of 67 students was designated the treatment group; students from the other five sections of the course made up the comparison group. The primary treatment manipulation was the type of test administered. Students in the treatment group were told that 60% of their course grade would be based on their performance on essay questions that required them to summarize or organize their knowledge. The remaining 40% of the grade would be based on their performance on multiple choice exams administered to all sections. The assumption here was, as Bloom, Madaus, and Hastings (1981, p. 39) put it, "For students, the objectives that really matter are those implicitly embedded in the tests on which their grades are based."

Subjects returned for testing 5 months after completing the course. Two of these measures showed interesting differences between treatment and comparison subjects. The first of these was a task in which students were asked to produce an outline of one of the main topics presented in the course. Figure 5 shows the results.

The dependent measure here is number of main ideas recalled. Treatment subjects who received high marks in the courses recalled significantly more main ideas than comparison subjects, whereas low achieving treatment subjects recalled less. Note that high achieving subjects in the comparison group did no better on this task than their low achieving colleagues.

The situation was completely reversed on the concept recognition task (see Figure 6). Here, comparison subjects performed better, especially high achievers and especially on the recognition of details (such as whether a particular illustration had been used in the text). Furthermore, Mittelholtz found that the best predictor of long-term retention

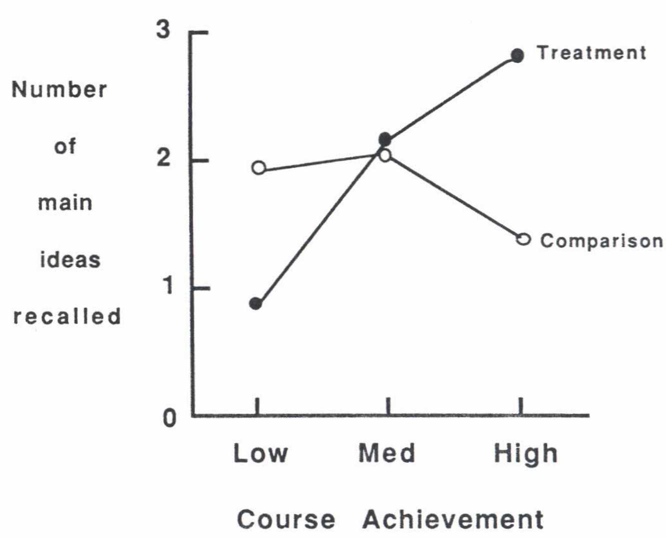

FIGURE 5. *Interaction between treatment and course achievement for number of main ideas recalled on the outline task (after Mittelholtz, 1988).*

for subjects in the treatment group was how well they had organized their essay responses during the course.

Finally, note that Mittelholtz looked at learning that occurred over an entire semester and that his most important findings concerned differences in retention some 5 months later. Much educational research looks at learning over one or two experimental sessions and ignores the delayed testing altogether. Thus, researchers give advice to educators that is certainly incomplete and may be misleading. This was shown in an aptitude-by-treatment-interaction study done some years ago by Yalow (1980). Yalow sought to investigate Galton's hypothesis that individuals differ in their ability to learn by visual or verbal elaboration and that the visualizers would do better in a visually rich treatment whereas verbalizers would do better in a verbally demanding treatment. This is what she found, at least on an immediate posttest. However, on a delayed posttest administered 1 month after the course, the interaction was reversed: Students did best when the treatment conflicted with their ability profile. One explanation for this is that learners do not get something for nothing. Materials that were easy to understand were also easy to forget. The low spatial student who had to struggle through a graph- and chart-laden treatment remembered what she had learned. In a similar way, much educational research is too myopic. We need to judge the efficacy of our efforts by their long-term effects, particularly their long-term transfer value. Unfortunately, the politics of higher education make such studies even less likely now than in the past. Researchers are rewarded—that is, tenured and promoted—for publishing lots of little studies. Few young enough to be enthusiastic about research have the luxury of waiting months or years to see how it all turned out.

Many different methods have been used in experimental studies to estimate how and how well learners have organized their knowledge (see Snow & Lohman, 1989, pp. 298ff). However, most of these methods are not suitable for classroom assessments. The Biggs and Collis (1982) scheme that Mittelholtz used has much to recommend it, although it, too, can be time-consuming to apply. Simpler procedures, such as asking students to produce an outline or to repair mistakes in one that is given to them, are much easier to use and, at least in Mittelholtz's (1988) study, showed high correlations with more laboriously obtained scores. However, such schemes can easily be abused. Educators who have heard the message about the importance of organization of knowledge have sometimes inadvertently imposed a common organizational structure. Students then end up memorizing outlines with the same misdirected zeal that they used to memorize details of text passages. Furthermore, a personally derived organization scheme is generally better than an externally provided scheme (Reese, 1977).

Achieving fluid organization. The ability to impose some organizational scheme on declarative knowledge is an important first step in acquiring a useful declarative knowledge base. Attempts to skip this step and move immediately to an instructional sequence that emphasizes multiple perspectives may overwhelm the beginning student. But there are many different ways to organize knowledge, especially in ill-structured domains. Such diversity can be encouraged by asking students to identify main points, but from several different perspectives. For example: What were some of the major events in U.S. history during the 19th century from the perspective of the federal government? From the perspective of Native Americans? From the perspective of Great Britain? Emphasizing main points, perhaps by integrating them into a story (Howard, 1991), but doing this from multiple perspectives may encourage the development of fluid abilities within and even across knowledge domains. Spiro et al. (1987) argue that the key is to structure instruction so that students develop multidimensional, multiply interconnected knowledge units that can be flexibly assembled in new ways when the student is confronted with ill-structured or far-transfer problems. They argue that

> The best way to . . . instruct in order to attain the goal of cognitive flexibility . . . is by a method of case-based presentations which treats a content domain as a landscape that is explored by "criss-crossing" it in many directions, by reexamining each case "site" in the varying contexts of different neighboring cases, and by using a variety of abstract dimensions for comparing cases. (p. 178)

Both Spiro et al. (1987) and Snow (1980) emphasize the importance of *assembling* situation-sensitive schema from knowledge units that are themselves fragmentary or are only "loosely coupled" (Snow, 1986). Spiro distinguishes between the prepackaged schema that are useful in solving problems in well-structured domains and the schema the problem solver must assemble anew to solve problems in ill-structured domains. Snow (1980) calls these crystallized and fluid abilities, respectively. Spiro et al. claim instructional methods that promote the orderly acquisition of the former may actually impede the development of the latter.[8]

The situation is summarized in Figure 7. As before, the abscissa represents amount of transfer. This time, though, the ordinate shows when the knowledge was learned. The area under the curve represents the potential range of transfer. Thus, recently acquired knowledge can usually be applied only locally. Old knowledge, on the other hand, can continue to be narrowly bound to a particular context or can be applied in an increasingly diverse array of contexts. Classroom quizzes typically require near transfer of recently acquired knowledge; far transfer is usually not yet possible.[9] Conventional tests of fluid abilities, on the other

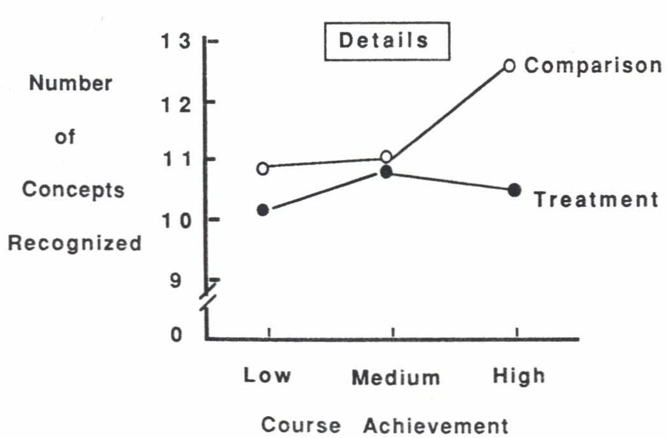

FIGURE 6. *Interaction between treatment and course achievement for the detail recognition task (after Mittelholtz, 1988).*

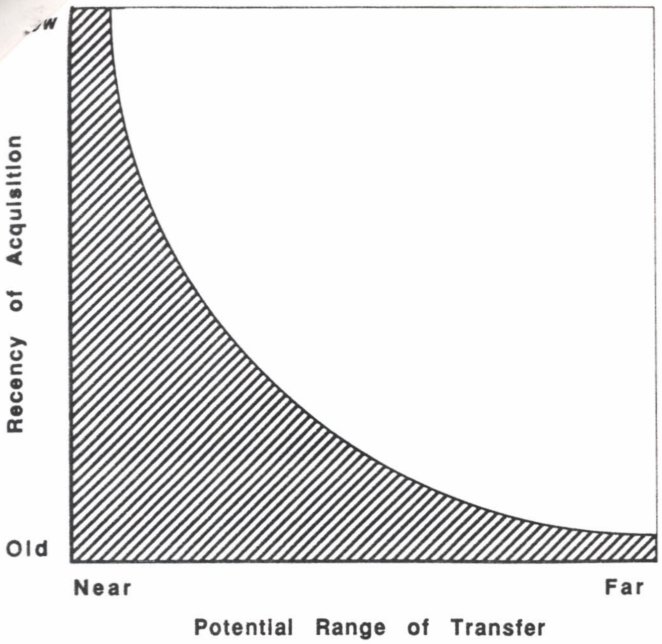

FIGURE 7. *Relationship between recency of knowledge acquisition and average potential range of transfer for that knowledge. More able subjects will show greater transfer than less able subjects.*

hand, use highly overlearned schema (such as knowledge of the alphabet or of geometric forms), but in problems each of which requires assembly of a new solution strategy. Figure 7 also suggests that transfer requires time to achieve. The fact that surveys consistently show that teachers rely primarily on knowledge-level questions may in part reflect the fact that this is what beginning learning looks like (what Rumelhart & Norman, 1976, call the accretion phase). Attempts to get teachers to create test items that demand more transfer might fare better if it were emphasized that the primary purpose of such questions is continually to stretch old knowledge in new ways.

Summary

In summary, my argument is that the ability to solve unfamiliar problems in a domain is one of the most important outcomes of formal schooling. These abilities are required in a general way on current tests of fluid ability. However, scores from existing tests of fluid abilities have few instructional implications. More instructionally useful tests can be devised by focusing on the extent to which particular abilities have been fluidized. This can be done by requiring students (a) to solve increasingly unfamiliar problems in a domain and (b) to impose multiple organizational schemes on their learning. In both cases, the motto might well be ''less is more.'' A clear understanding of the key events, controversies, or concepts in a domain, along with the ability to connect these ideas both to each other and to a larger scheme is more important than a much larger base of fact and skill knowledge that is disconnected, is not tied with other learning, and can be applied only locally. Further, effects seem to be magnified by student ability and educational level. For example, Mittelholtz (1988) found that the effects of tests that emphasized organization and transfer were greatest for the most able students in college courses. Similarly, I would expect that the importance of a rich but flexibly organized knowledge base would increase as the amount of knowledge represented in the system increases. I do not need a card catalog for the books on the shelves in my office, but the university libraries would be useless without one. In like manner, a flexibly organized knowledge base may be more important for the development of expertise than for the attainment of minimum competence. Finally, I have argued that it is time to exorcise the notion that good assessments will always yield scores that permit individuals to be ranked, on one or several scales. Although teachers must emphasize organization and transfer in both their teaching and their testing, the fact that there are multiple organizational schemes and different perceptions of task novelty makes it impossible to devise tests that clearly rank order test takers on their ability to transfer or to organize. Like the instructional methods that encourage the development of these abilities, instructionally useful tests of fluid abilities demand serious consideration of the learner's perspective, and not simply the common perspective codified in the curriculum.

Notes

This article is based on an invited address to the Henry B. and Jocelyn Wallace National Research Symposium on Talent Development in May 1991. I am indebted to Julian Stanley, Nancy Jackson, and other conference participants who encouraged me to seek a wider audience for these remarks.

[1]Although many have made this assertion, it is unlikely that it will be accepted until new measures of intelligence can be devised that show high correlations with old measures but that are also instructionally tractable.

[2]One can advocate an investment theory of aptitude without assuming that fluid abilities are innate (e.g., Horn, 1985).

[3]Contradictions make it difficult to summarize this literature. It appears that intelligence tests show higher heritabilities when the sample is composed of children (rather than of adolescents or adults) who vary widely in age and when the tests of fluid ability emphasize spatial reasoning and working memory, or both (as in tests of inductive reasoning). This may reflect the fact that although performance on spatial reasoning and working memory tasks can be improved with experience and practice, these abilities are not directly trained through conventional school activities and thus show less differential growth. It may also reflect something peculiar in the estimation of heritability for achievement using multilevel achievement batteries in age-heterogeneous samples, since heritability coefficients for achievement in these studies are often surprisingly low (e.g., less than .20 in Thompson et al., 1991).

[4]Recently, some have once again argued that transfer is an improbable goal of well-intentioned but poorly informed educators. Thorndike himself (1913) offered perhaps the most forceful rejection of this view, and argued that even general transfer was commonly observed. Two facts need emphasizing: First, it is useful to distinguish between a general-to-specific dimension of transfer (the complement to Gagne's, 1970, vertical transfer) and a near-to-far dimension of transfer (Gagne's, 1970, lateral transfer). Second, in comparisons of near and far transfer, by definition, near transfer will be easier to achieve and easier to measure. This means that horse races between the two will invariably be won by near transfer, which makes horse races uninformative. This is because the number of situations that constitute the domain of transfer increases with the amount of transfer required. Experiments (or anecdotes) that compare the probability of near versus far transfer by comparing performance on one task judged to require near transfer and another judged to require far transfer must adjust for these different probabilities. To my knowledge, no one has attempted to do this. The problem is further complicated by the fact that a common measuring stick becomes increasingly undependable as average transfer distance increases.

[5]Multiple choice tests are often improperly characterized as testing "only recall." As I see it, such tests often emphasize recognition instead.

[6]A year later, I asked a child from this class if she remembered this question. She said, "Oh, yes that was Thomas Jefferson." I then asked her if she knew anything else about him. She thought and said, no, not that she could remember and then, "Oh yes, wasn't he a president or something?"

[7]Content analyses of classroom tests (Fleming & Chambers, 1983) support these claims, although generalization is difficult since items are invariably classified using the Bloom (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) taxonomy. This national model has no category for assessment of the organization of factual knowledge.

[8]There are interesting parallels between the approach of Spiro et al. (1987) and Langer's (1989) concept of mindfulness. Mindful individuals are said to be open to the possibility of viewing situations in novel ways, are sensitive to context, and routinely invent new categories to order experiences.

[9]Sometimes newly acquired knowledge can be more broadly applied, especially when there is a deliberate effort to discover far transfer (Salomon & Perkins, 1989). See also note 4.

References

Anastasi, A. (1937). *Differential psychology*. New York: Macmillan.

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

Angoff, W. H. (1988). The nature-nurture debate, aptitudes, and group differences. *American Psychologist, 43*, 713–720.

Balke-Aurell, G. (1982). *Changes in ability as related to educational and occupational experience* (Göteborg Studies in Educational Sciences, No. 40). Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives handbook: The cognitive domain*. New York: McKay.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. New York: McGraw-Hill.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego: Academic Press.

Burt, C. (1955). *Mental and scholastic tests* (4th ed.). London: Staples.

Butterfield, E. C., & Nelson, G. D. (1989). Theory and practice of teaching for transfer. *Educational Technology Research and Development, 37*(3), 5–38.

Campione, J. C., Brown, A. L., & Ferrara, R. A. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 392–490). New York: Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

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

Cattell, R. B. (1982). *The inheritance of personality and ability: Research methods and findings*. New York: Academic Press.

Ceci, S. J. (1990). *On intelligence . . . more or less*. Englewood Cliffs, NJ: Prentice Hall.

Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics, 48*, 1074–1079.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1–14.

Dahlgren, D. O. (1979). *Understanding students' understanding: Some qualitative aspects of process and outcome of teaching and learning at university level* (Report from the Institute of Education, No. 80). Göteborg, Sweden: University of Göteborg.

Davis, B. L., & Flaherty, P. (Eds.). (1976). *Human diversity: Its causes and social significance*. Cambridge, MA: Ballinger.

Elshout, J. J. (1983). Is measuring intelligence still useful? In S. B. Anderson & J. S. Helmick (Eds.), *On educational testing* (pp. 45–56). San Francisco: Jossey-Bass.

Eysenck, H. J. (1982). *A model for intelligence*. New York: Springer.

Fleming,, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. *New Directions for Testing and Measurement, 19*, 39–49.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*, 193–202.

Freeman, F. S. (1934). *Individual differences*. New York: Holland.

Gagne, R. M. (1970). *The conditions of learning* (2nd ed.). New York: Holt, Rinehart & Winston.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Glaser, R. (1972). Individuals and learning: The new aptitudes. *Educational Researcher, 1*(6), 5–12.

Halpern, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Erlbaum.

Heidegger, M. (1982). *The basic problems of phenomenology*. Bloomington, IN: Indiana University Press.

Hieronymus, A. N., Hoover, H. D., & Lindquist, E. F. (1986). *Iowa Tests of Basic Skills* (Forms G/H). Chicago: Riverside.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1987). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research on Aging, 3*, 33–84.

Howard, G. S. (1991). Culture tales: A narrative approach to thinking, cross-cultural psychology, and psychotherapy. *American Psychologist, 46*, 187–197.

Humphreys, L. G. (1981). The primary mental ability. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 87–102). New York: Plenum.

Humphreys, L. G. (1986). Describing the elephant. In R. J. Sternberg & D. K. Detteman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 97–100). Norwood, NJ: Ablex.

Hunt, J. M. (1961). *Intelligence and experience*. New York: Ronald.

Janey, T. (1991). *Measuring transfer in an academic setting*. Unpublished doctoral dissertation, The University of Iowa, Iowa City, IA.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Kennedy, M. M. (1978). Findings from the Follow Through Planned Variation Study. *Educational Researcher, 7*(6), 3–11.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning is (little more than) working memory capacity. *Intelligence, 14*, 389–433.

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 41–71). Hillsdale, NJ: Erlbaum.

Lynn R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences, 11*, 273–285.

Marton, F. (1983). Beyond individual differences. *Educational Psychology, 3*, 291–305.

Marton, F., Svensson, L. (1979). Conceptions of research in student learning. *Higher Education, 8*, 471–486.

Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19–40). Hillsdale, NJ: Erlbaum.

Mittelholtz, D. J. (1988). *The effects of essay tests on long-term retention of course material*. Unpublished doctoral dissertation, The University of Iowa, Iowa City, IA.

Norman, D.A. (1977). Notes toward a theory of complex learning. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 29-48). New York: Plenum.

Ottosson, T. (1987). *Map reading and way finding*. (Göteborg Studies in Educational Sciences, No. 65). Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 73–98). Hillsdale, NJ: Erlbaum.

Raven, J. C. (1962). *Advanced Progressive Matrices*. New York: Psychological Corporation.

Reese, H. W. (1977). Imagery and associative memory. In R. V. Kail

& J. W. Hager (Eds.), *Perspectives on the development of memory and cognition* (pp. 113–176). Hillsdale, NJ: Erlbaum.
Rice, T., Fulker, D. W., DeFries, J. C., & Plomin, R. (1988). Path analysis of IQ during infancy and early childhood and the index of the home environment in the Colorado Adoption Project. *Intelligence, 12,* 27–45.
Rumelhart, D. E., & Norman, D. A. (1976). *Accretion, tuning, and restructuring: Three modes of learning* (Rep. No. 7602). San Diego: University of California, Center for Human Information Processing.
Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 36,* 115–140.
Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam, Netherlands; Elsevier North-Holland.
Scarr, S., & Carter-Saltzman, L. (1982). Genetics and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 792–896). Cambridge, MA: Cambridge University Press.
Sharp, D., Cole, M., & Lave, C. (1979). Education and cognitive development: The evidence from experimental research. *Monographs of the Society for Research in Child Development, 44,* Nos. 1-2.
Sherman, M., & Key, C. B. (1932). The intelligence of isolated mountain children. *Child Development, 3,* 279–290.
Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.
Snow, R. E. (1978). Theory and method of research on aptitude processes. *Intelligence, 2,* 225–278.
Snow, R. E. (1980). Aptitude and achievement. *New Directions for Testing and Measurement, 5,* 39–59.
Snow, R. E. (1982). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493–585). Cambridge, England: Cambridge University Press.
Snow, R. E. (1986). On intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and development* (pp. 133–140). Norwood, NJ: Ablex.
Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: Macmillan.
Spiro, R. J., Vispoel, W. P., Schmitz, J. G., Samarapungavan, A., & Boerger, A. E. (1987). Knowledge acquisition: Cognitive flexibility and transfer in complex content domains. In B. K. Britton & S. W. Glynn (Eds.), *Executive control processes in reading* (pp. 177–199). Hillsdale, NJ: Erlbaum.
Stankov, L. (1986). Kvashchev's experiment: Can we boost intelligence? *Intelligence, 10,* 209–230.
Stern, W. (1914). *The psychological methods of testing intelligence* (Educational Psychology Monographs, No. 13). (G. M. Whipple, Trans.). Baltimore: Warwick & York.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge, MA: Cambridge University Press.
Sternberg, R. J., & Gastel, J. (1989). Coping with novelty in human intelligence: An empirical investigation. *Intelligence, 13,* 187–197.
Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.
Terman, L. M. (Ed.). (1925). *Genetic studies of genius. Vol. I. Mental and physical traits of a thousand gifted children.* Stanford, CA: Stanford University Press.
Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science, 2,* 158–165.
Thorndike, E. L. (1913). *Educational psychology: Vol. 2. The psychology of learning.* New York: Teachers College, Columbia University.
Thorndike, R. L. (1963). *The concepts of over- and under-achievement.* New York: Columbia University, Teachers College.
Thorndike, R. L., & Hagen, E. (1986). *Cognitive Abilities Test* (Form 4). Chicago: Riverside.
Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247–261.
Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1.*
Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Walkins.
Yalow, E. S. (1980). *Individual differences in learning from verbal and figural materials* (Tech. Rep. No. 12). Stanford, CA: Stanford University, School of Education, Aptitude Research Project (NTIS No. AD-A098431).

1995 Annual Meeting Program Theme

The theme of the 1995 Annual Meeting of the American Educational Research Association will be Partnerships for a New America (tentative title). The theme is grounded in growing acceptance of the need for interdependence among the human service professions (e.g., education, social work, school psychology, public health administration) in order to effectively serve children, adults, and families in America. It is apparent that partnerships of this nature will transform schools and their relationships with families and community agencies. These evolving School-University-Community partnerships will alter many of the strategies currently in practice and require new ones. There are crucial questions regarding the knowledge base associated with these changes and partnerships. The 1995 AERA Annual Meeting will focus on pioneering research and scholarly efforts associated with interdisciplinary partnerships and the resulting interprofessional collaboration.

So that the research is guided and focused in this area, Jane Stallings, AERA president-elect is announcing the theme 18 months before the Annual Meeting allowing researchers to develop cross-disciplinary research agendas. There is need for data-based research and theoretical frameworks relevant to these new partnerships, and their effects upon school and college faculty, new professional training, children, families, schools, and communities. Case studies based on multiple perspectives are welcomed, as well as experimentally designed programs. If data are collected now, researchers will be ready to submit their proposals in the 1994 August Call for Papers. Please contact Jane Stallings, Dean of Education, College of Education, Texas A&M University, College Station, TX 77843-4222, with ideas and information.