Svenson, O., & Hedenborg, M.L. (1979). Strategies used by children when solving simple subtractions. *Acta Psychologica, 43,* 477-489.

Van Lehn, K. (1983). Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. *Journal of Mathematical Behavior, 3,* 3-71.

Winkelman, H.J., & Schmidt, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology, 102,* 734-736.

Woods, S.S., Resnick, L.B., & Groen, G.J. (1975). An experimental test of five models for subtraction. *Journal of Educational Psychology, 67,* 17-21.

Young, R.M., & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science, 5,* 153-177.

Shepard, L.A. (1992). In B.R. Gifford & M.C. O'Connor (Eds.) Changing Assessments: Alternative Views of Aptitude, Achievement, and Instruction. Boston: Kluwer Academic Publishers. www.kluweronline.com

Commentary:
What Policy Makers Who Mandate Tests Should Know About the New Psychology of Intellectual Ability and Learning

Lorrie A. Shepard

This volume addresses itself to policy makers but is inaccessible to them. These chapters, written by some of the most prominent researchers in contemporary psychology, are long, dense, and complex. Can it be imagined that a state legislator, concerned about highway funding one day and re-election the next, would sit down some evening and digest the contents of this book? Does one imagine, even, that school board members whose purview is focussed on educational issues will read and reread these pages so as to reframe their conceptions of learning and assessment? And why should policy makers want to know about psychological theory or what psychologists think about assessment? Because, many educational policy decisions about categorical programs for handicapped children, compensatory education, grade-to-grade promotion standards, or mandated accountability tests are based implicitly on policy makers' own "theories" about what conditions of education will foster student learning. If they are unaware of new research findings about how children learn, policy makers are apt to rely on their own implicit theories which most probably were shaped by the theories that were current when they themselves attended school. Scientific knowledge about the development of intellectual ability and learning is vastly different today than what was known 40 or 50 years ago. Some things that psychologists can prove today even contradict the popular wisdom of several decades ago. Therefore, if policy makers proceed to implement outmoded theories or tests based on old theories, they might actually subvert their intended goal--of providing a rigorous and high quality education for all students.

The purpose of this commentary is to translate for policy makers the most important findings from the "new psychology" of intelligence and learning, and to summarize the implications of our current knowledge for assessment practices. The commentary is organized into three parts. In the first section are summaries and critiques of each chapter. In the second section consensus views are presented, summarizing across the chapters what cognitive science tells us today about the nature of intelligence and the nature of learning and achievement.

Finally, implications of the new psychology are considered for three assessment applications: classroom assessment, assessment for special placement and tracking, and accountability assessment.

CHAPTER SUMMARIES AND CRITIQUES

Each chapter summary is intended to provide an overview of important ideas and explanations for lay audiences. Because my aim is to arrive at consensus views, presented in the next major section of the commentary, I tend to give uneven treatment to key ideas in this section. Here I elaborate on unique aspects of each author's contribution and discuss my criticism of ideas that are excluded from the later consensus. The most important ideas, which form the basis of the later consensus, are mentioned here but are discussed in greater detail in later sections of the commentary.

The Resnicks' Chapter: Assessing the Thinking Curriculum: New Tools for Educational Reform

Lauren Resnick is an internationally recognized cognitive researcher whose work bridges the worlds of psychology and education. She is a past president of the American Educational Research Association and the founding editor of the journal, Cognition and Instruction. Daniel Resnick is an eminent historian of American education, with particular expertise in the history and social functions of testing. Thus Daniel Resnick broadens the perspective of the volume that is otherwise focused on cognitive psychology.

More than any of the other authors, the Resnicks address themselves to policy makers. Authors of the remaining chapters consider how assessment should be transformed based on current theories of learning and ability, but they are interested in making accurate decisions for individual test takers primarily for instructional purposes. The Resnicks talk directly to policy makers about accountability assessment, i.e., the kinds of tests that policy makers are responsible for. What kinds of measures should be given to groups of students to judge, in the aggregate, the quality of educational programs? They help policy makers see how the character of the mandatory, external assessments they choose will foster or hinder educational reform.

The Resnicks focus on the goal of education which is to teach students to think. Much of current educational practice derives from the mass-education system of the past century which was designed to teach rudimentary basic skills to the majority, reserving higher-order intellectual pursuits for an elite few. Today, educational reformers recognize that the current economic environment demands that all students, not just the elite, be taught to reason, adapt, and solve problems in an ever-changing work environment.

Unfortunately the mass-education system's emphasis on the basics is ill suited to helping students develop the ability to think. The most important

contribution in the chapter and in the volume is the insight from contemporary research that all learning involves thinking. It is incorrect to believe, according to old learning theory, that the basics can be taught by rote followed by thinking and reasoning. As documented by the Resnicks, even comprehension of simple texts requires a process of inferring and thinking about what the text means. Children who are drilled on number facts, algorithms, decoding skills, or vocabulary lists without developing a conceptual model or seeing the meaning of what they are doing have a very difficult time retaining information (because all the bits are disconnected) and are unable to apply what they have memorized (because it makes no sense). These ideas from current learning theory are developed further in the consensus section.

Having described the kind of curriculum that must exist to foster thinking, the Resnicks turn to current testing practices which are inimical to the goal of teaching thinking. Key assumptions from out-moded learning theory, which the Resnicks refer to as the decomposability and decontextualization assumptions, are carried forward by present-day tests. Psychological theories of the 1920s assumed that learning of complex competencies could be broken down into constituent skills (into individual stimulus-response bonds) and learned one component at a time. The Resnicks use the metaphor of a machine that is put together after all the component parts have been manufactured. The old theory never specified, however, how the parts were to be assembled into thinking. Similarly according to the decontextualization assumption component skills are generic and do not depend on the context in which they are learned or applied.

One of the most useful and compelling parts of the Resnicks' chapter is their analysis of specific standardized tests in light of contemporary learning theory. The best multiple-choice tests are reading comprehension tests which nonetheless have students reading short passages and searching for right answers at a rate of one-per minute. Multiple-choice questions elicit superficial comprehension rather than deep reflection. (Given the rules of the test-makers art, they certainly never invite students to wrestle with ambiguity.) The nature of the test tasks is much worse in other skill areas, requiring recognition of errors and computational fluency rather than thinking. The character of test questions is particularly worrisome if one begins to ask not just how well the items measure what students know, but how well they serve as templates for instruction. "Children who practice reading mainly in the form in which it appears on the tests--and there is good evidence that this is what happens in many classrooms--would have little exposure to the demands and reasoning possibilities of the thinking curriculum." (p. 46) And again, "Students who practiced mathematics in the form found in the standardized tests would never be exposed to the kind of mathematical thinking sought by all who are concerned with reforming mathematics education..." (p. 47) (see also National Council of Teachers of Mathematics, 1988; National Research Council, 1989).

Next the Resnicks offer an analysis of the different functions of testing as a framework for their subsequent in-depth treatment of accountability assessment. The contrasts between what is needed for instructional tests and for accountability tests are taken up in later sections of the commentary. For example, instructional assessments must be much more frequent than accountability measures but do not require disinterested verification of teacher judgments.

Accountability assessment exerts a potent influence on school curriculum because of the importance attached to publicly reported test results. Tests are not unobtrusive measures like thermometers; they influence how educators behave. Many testing programs have been instituted intentionally to leverage school reform. What the Resnicks' analysis reveals, however, is that "measurement-driven instruction" will lead reform in the wrong direction if tests embody incomplete or low-level learning goals. They deduce three guiding principles for accountability assessments:

> 1. You get what you assess. "Educators will teach to tests if the tests matter in their own or their students' lives" (p.59).
>
> 2. You do not get what you do not assess. "What does not appear on tests tends to disappear from classrooms in time" (p.59).
>
> 3. Build assessments toward which you want educators to teach. "Assessments must be designed so that when teachers do the natural thing--that is, prepare their students to perform well--they will exercise the kinds of abilities and develop the kinds of skill and knowledge that are the real goals of educational reform" (p.59).

In the last section of the commentary I elaborate on the Resnicks' recommendations for performance assessments. In essence they have laid out what policy makers should do, in the name of accountability, if they are serious about reforming schools so that all students will be taught to think.

Gardner's Chapter: Assessment in Context: The Alternative to Standardized Testing

Howard Gardner is a famous Harvard psychologist who has popularized the concept of multiple intelligences. His chapter begins with a stark portrayal of the excessive reliance today on formal testing. Current testing practices derive both from out-moded theories about human capacities and the cult of efficiency in U.S. society. Gardner then offers explicitly a summary of contemporary scientific knowledge that should serve more appropriately as the basis for new approaches to assessment. Key points, especially the variety of intelligences and the influence of context, are discussed in a later section of this chapter regarding the nature of intelligence.

The spirit of Gardner's recommendations for alternative assessments is captured by his envisioning of an apprenticeship model. The novice should be evaluated from the all-knowing, teaching perspective of the master rather than by a remote and decontextualized test. New approaches to assessment should be a natural part of the learning environment and should preserve "ecological validity." Assessments should tap the full range of intelligences, using multiple measures. Assessors should be responsive to individual differences and use materials that are intrinsically interesting and motivating. The purpose of assessment should be to help individual students.

Gardner then goes on to describe in considerable detail two assessment projects based on these principles. The first is a preschool program with activities designed to measure a broad array of cognitive strengths as well as children's styles in approaching various tasks. The second is a secondary school arts project intended to assess intellectual competence in creative writing, graphic arts, and musical performance. The final portion of the chapter is devoted to Gardner's conception of the ideal school which would feature assessment specialists, student-curriculum brokers, and school-community brokers.

In my opinion, the details of Gardner's assessment examples provide two powerful insights: 1. It is possible, although exceedingly difficult, to implement complex, broad, performance-based assessments in real learning environments. (I return to Gardner's development of domain-projects and process-folios of students' work later, in the context of discussing "authentic" assessments for educational accountability.) 2. It is impossible to assess aptitudes for as yet undeveloped intelligences without providing instruction. Thus, Gardner and his associates found they could not maintain a distinction between developing assessment and developing curriculum. In the preschool setting, "our approach has been to expose students to rich experiences in the particular domain of interest and to observe the way in which they become engaged in that domain. The ensuing record provides a powerful indication of how much talent or potential such students exhibit in the domain of interest" (p.96). In the arts project, students are provided with instruction and exercises. Assessment of students' talent is based as much on progress in response to instruction as on the technical quality of finished products. These examples, where students do not have to have had enriched prior experiences to fare well on the assessments, are more compelling to the reader than merely asserting the conclusion of contemporary psychology that "IQ is developed." Furthermore, they convey concretely how extensive efforts must be really to provide opportunities to develop an array of aptitudes.

Having praised Gardner's insight, I nonetheless have the following criticisms of his chapter: 1. He is too quick to turn brilliant and promising research into practice, without validity evidence and without an appreciation of negative side-effects in real school contexts. 2. Gardner's very broad conception of intelligence unwittingly preserves a very narrow and elite view of the two

academic intelligences. 3. His ideal school with separate assessment and curriculum specialists recreates all the mistakes of Special Education, especially it errs by separating assessment from instruction and disempowering teachers. In the following paragraphs I explain these points briefly.

Gardner himself mentions the potential risk of early labelling. Nonetheless he is willing to send preschool profiles home to parents with the caveat that a child's strengths and weaknesses may not be stable over time. Gardner's assessment exercises are highly experimental. He does not know whether they are reliable or predictive of adult profiles. He has not examined whether children's apparent talents and interests are "real" (i.e., hold true across all of the tasks in a given domain) or depend on idiosyncratic features of selected activities. It is not known whether the game-like activities are differentially attractive to children because of race or gender stereotypes. The effect of training strengths or of training weaknesses following assessment has not been investigated systematically. Given a history where lay persons are likely to over-interpret "scientifically" derived test scores, and a climate where "yuppie parents" may well feel compelled to instruct their child's strengths or weaknesses, it seems irresponsible to release the results of preschoolers' profiles. The entire history of intelligence testing, which Gardner recounts derisively, might have been different if researchers making the first tentative efforts to understand human intelligence had not been so eager to turn their fallible measuring devices into practical applications like army selection and school placement tests.

A fundamental question regarding Gardner's work is the appropriateness and timing of specialization. If a child has potential talent to be a dancer, at what time should development of "bodily-kinesthetic intelligence" be at the expense of the other intelligences? Because Gardner's contribution has been to focus on talents missed by traditional schools and tests, he tends to ignore the possibility explained by the Resnicks that academic intelligences have been underdeveloped by opportunities in schools, especially for some groups of students. Consider the following statements:

> In truth, I do not worry about those students who are excellent in linguistic and logical pursuits. They will likely find their rewards within the school, in standard gifted programs, or in special advanced sections or honors group. The educational challenge is to provide correlative kinds of opportunities for students who have cognitive and personal strengths which, however, are not well addressed by the standard curriculum in school (p.111).
>
> I believe that, especially when resources are scarce, every individual ought to have the opportunity to show his or her strength. There is no objection to a "high scorer" being able to show off his string of 800s to a college admissions staff; by the same token, individuals with other cognitive or stylistic strengths ought to have their day as well (p.114).

Although Gardner's own analysis can be used to argue that there are individuals with potential strengths in linguistic and logical-mathematical intelligences who do not do well on decontextualized tests or in school, he proceeds as if individuals with strengths in these areas were well assessed and well served. In contrast to the Resnicks who propose more rigorous academic opportunities for all students, there is the dangerous inclination in Gardner's work to give up on those who are not already accomplished academically and focus on "vocational training" (albeit of an intellectual sort).

Gardner's ideal for the future is an individual-centered school that would differentiate instruction according to individual assessed strengths and weaknesses. While Gardner's motive is to develop talents that have hitherto been ignored, there is reason to fear, given a realistic understanding of the sociology of schools, that these ideas could foster tracking and denial of opportunities. Curiously, the extent of differentiation among students that Gardner has in mind causes him to propose separate assessment specialists and curriculum brokers as well, leaving teachers free to teach subject matter. He thus abandons his own apprenticeship model where the master-teacher knows intimately the development and learning style of the apprentice. His own assessment projects suggest that the observer learns more from interacting with students than can be summarized in a set of scores, yet he proposes to separate the roles of teacher and assessor (presumably because students will be tracked into such different curricula that they cannot be served by a common set of teachers). Special Education is replete with examples of what happens when valid clinical assessment procedures are institutionalized and bureaucratized. Not only are the resulting student placements often inaccurate by scientific criteria, but costly assessment procedures are reported by teachers to serve no instructional purpose except to justify placement (Shepard, 1983). Gardner's school of the future is hypothetical, based on many idealized assumptions, including the assumptions that a huge cadre of highly trained assessors would be available, that assessments would be sufficiently accurate not to risk misclassifying students on the basis of past opportunities, and that teachers and assessors would always act in the best interests of students rather than the efficiency of the school. If one is imagining an ideal, however, why not imagine teachers with the training of Gardner's assessors and a system of instruction that is so effective in developing all the intelligences in all individuals that there is no need to differentiate for specialization until very late in individuals' school careers?

Brown, Campione, Webber, and McGilly's Chapter: Interactive Learning Environments: A New Look at Assessment and Instruction

Ann Brown and Joseph Campione are two of the most eminent cognitive researchers in the United States today. Although perhaps unknown to policy makers, they and their colleagues are highly esteemed in the academic world for

their exhaustive research on the development of intelligent thought processes--e.g., how learning, remembering, and understanding occur. Unlike most laboratory scientists, they have managed to make significant contributions to both psychological theory and educational practice. For example, based on what is known about how higher-order thinking processes are developed, Palincsar and Brown (1984) were able to improve remarkably the reading comprehension strategies of poor junior-high-school readers who had not been helped up to that time by years of remedial instruction.

Brown, Campione, Webber, and McGilly's chapter is the individual assessment companion to the Resnicks' accountability assessment chapter. The two chapters share a common view of learning and the type of instructional environment essential to teaching children to think. What Brown et al. provide is a detailed picture of the kinds of instructional efforts that could ensue in classrooms if teachers had to answer to the Resnicks' kind of accountability measures rather than to standardized tests. Brown et al. are interested in developing dynamic assessment techniques, closely tied to instruction, that would help a teacher see how a student is learning and what new understandings and insights are just next within the student's reach. As can be seen in the structure of the chapter, the chronology of their work has two distinct phases. Early on, Brown and Campione were interested in assessment procedures that would permit a more accurate determination of learning potential. Insights from these studies have moved them more and more to focus on instructional methods that actually change children's ability to learn.

Brown et al. begin with a history of the concept of intelligence. There is no better way to see the import of current research understandings than to contrast them with old beliefs. In the past, IQ was seen as a highly stable and generalized trait. This view made it reasonable to believe that static, one-time tests could locate students accurately on a continuum of teachability. Accordingly tests were used to make permanent placement and selection decisions, in keeping with the student's relative standing, rather than attempting to disrupt or alter the student's existing capacity to learn. Problems that persist today in tracking and labelling students can be linked back to traditional theories of intelligence.

Next Brown et al. summarize the seminal theories of Vygotsky and Feuerstein--in a way very helpful to the uninitiated reader. Both theorists saw intellectual abilities as malleable and explained the development of cognitive processes in terms of highly interactive, social experiences. Both were clinicians dealing with individuals from severely deprived learning environments, Vygotsky with children raised after the dislocation of the Russian Revolution, and Feuerstein with refugees from World War II. If individuals have had markedly limited opportunities to learn, it is clearly not possible to assess from what they know what their capacity to learn might be. Therefore, both Vygotsky and Feuerstein, like Brown and Campione after them, developed assessment techniques based on focused intervention so that they could learn something from how students responded to instruction. If static assessment is inaccurate, better to draw conclusions about capacity to benefit from instruction by staging real instructional experiments. It is impossible to distinguish whether each of these theories is a developmental theory, an assessment model, or an instructional paradigm--all are entwined. If the reader has never heard of Vygotsky, this portion of the Brown et al. chapter deserves careful attention. Presently there is an enormous resurgent interest in Vygotskian theory--focused on the social construction of meaning--affecting research and reform in every curricular area from reading comprehension to science concepts.

Guided assessment is the term Brown et al. use to refer to techniques based on observing students in the process of learning. The most lengthy section of the chapter is devoted first to describing a taxonomy of approaches to guided assessment followed by a review of the major programs of research and clinical intervention programs. All but the most diligent students and scholars will be tempted to skip this section. However, it is worthwhile for the reader to pick any one of the research programs and follow the arguments and evidence presented in detail. Understanding the kinds of investigations undertaken and their extent helps one to appreciate that assertions about helping students to "learn to learn" are not mere rhetoric. Evidence to support what Brown et al. call the corrigibility of intelligence is substantial. Furthermore, when guided assessment approaches are compared to traditional measures, it is demonstrated repeatedly , as often claimed by political groups, that traditional measures dramatically underestimate the learning potential of some children.

The authors' own version of guided assessment, called "dynamic assessment," is based on Vygotskian theory. The assessment-teaching effort begins with a pretest of what the child already knows. The assessor-teacher has to judge where the child is ready to begin. Then the child and teacher work cooperatively with the teacher providing increasingly more specific hints until the child learns to solve the assessment problems of a certain type independently. (Measurement of the child's ability is based on the number and specificity of hints required.) The process continues with the teacher providing support and hints to aid transfer of the student's learning to more and more different problem types. Thus children learn in the course of being assessed. After their initial studies aimed at assessing-teaching general inductive reasoning skills, Brown and her colleagues directed their efforts toward assessing thinking skills in subject-matter contexts. They did this to ensure that assessment would contribute directly to the instructional process.

The great promise of dynamic assessment poses a considerable practical dilemma, however. Because the procedures require labor-intensive protocols administered by well-trained assessors, the demands of the program seem to be nothing less than full-time one-on-one tutoring by expert teachers. At different points in their discussion, Brown et al. consider three possible resolutions to the

dilemma: assess only children with seriously learning difficulties (the Special Education model), develop intelligent computer programs to conduct the assessments, or develop instructional programs where the assessment steps are implicit.

Along with many other cognitive psychologists, Brown et al. see the computer as a means to help teachers with the impossible logistical demands of individual diagnostic assessment. If it were possible to develop a formal model of expert understanding of a subject area and to model developmental stages as well as typical missteps in acquiring mastery, then the computer could be used to conduct dynamic assessments, providing hints and giving instantaneous feedback. Examples of such intelligent tutors have already been tried out in very circumscribed instructional domains; for example, Brown et al. use the example of teaching place value. Perhaps these endeavors hold promise for the distant future. However, the cognitive mapping of most subject matter areas is still too primative to lend itself to computerized teaching. Furthermore, even in the future, it remains to be seen whether the efficiency and immediacy of computerized tutoring can substitute for the social aspects of meaning-making that students gain when they interact with teachers and peers. Reliance on computers is the only part of Brown et al.'s brave-new-world image of the future that I find disquieting. The branch in their work toward computers seems not to have preserved the integrity of Vygotskian social-constructivism as well as the path toward instructional paradigms, which we consider next. In any case, an appropriate way to evaluated the effectiveness of computerized assessments would be to compare their cost against an equal investment spent to train teachers to think about student thinking.

In the penultimate section of the chapter, Brown et al. make the shift from interactive assessment (of learning potential) to instruction. Although Brown and her colleagues embark on this course so as to integrate assessment and instruction, the reader will note that the transformation in their work in really more profound that this. As they undertake real and extensive instructional programs, assessment disappears. The Reciprocal Teaching model provides a format for instruction so that students will have extensive guided practice with the kinds of thinking skills that are most valued. In the case of reading comprehension, students are trained explicitly to use strategies of questioning, summarizing, clarifying, and predicting, when discussing the meaning of a text they have read with their peers. In mathematics, students are explicitly taught the strategies of problem identification, assembling relevant information, problem representation and problem solving, and checking. The instructional programs that Brown et al. describe share several features derived from cognitive-constructivist research. In each application, the reasoning skills that are taught explicitly are those that good students tend to develop implicitly. The instructional activities are interactive and egalitarian providing for social, co-construction of meaning. And, in keeping with the Resnicks' and Gardner's

arguments against decontextualized skill development, their teaching activities preserve the integrity of the target task by handling component skills in the context of larger tasks. Their instructional programs are aimed at developing higher-order thinking in subject matter domains because reasoning ability depends on domain-specific knowledge and because it is indefensible to divert so much instructional time unless the content is useful.

In the end, Brown et al. give very little advice to policy makers because their interest is the assessment and guided instruction of individual learners. Stated or unstated, however, the insights in their chapter have tremendous policy implications. First, it should be clear that the kind of assessment that will genuinely improve student learning and thinking cannot be conducted from the statehouse. Second, large-scale implementation of the kind of instruction that Brown et al. have described would make it increasingly unnecessary to select and assign students to a separate place for special education or remedial help.

Sternberg's Chapter: CAT: A Program for Comprehensive Abilities Testing

Robert Sternberg is a Yale psychologist famous for his research on intelligence. As can be seen from his reference list, he has conducted empirical studies on components of intelligence, but he is best known for his efforts to produce a grand, synthesizing framework to conceptualize all aspects of mental ability.

In the first part of his chapter, Sternberg recapitulates the history of conflicting views of intelligence, and presents his integrative theory--called the triarchic theory because of its tripartite structure. In addition to internal mental processes, such as knowledge-acquisition components and the ability to monitor one's own problem solving efforts, Sternberg's theory acknowledges the effects of prior experience and contextual demands on an individual's manifest intelligence.

Sternberg's triarchic theory exemplifies contemporary psychological thought about the nature of intelligence. Especially, he explains that intelligence is multidimensional (thus, individuals are likely to have very different profiles of strengths and weaknesses) and hierarchical (higher-order processes direct the use of lower-order processes). Mental ability is also developed, rather than fixed, and the use of intelligence is dependent on context. It is these themes in Sternberg's work that I rely on to develop the consensus view of ability in the discussion that follows.

After the triarchic theory, Sternberg's chapter becomes problematic. He offers a new test to measure intellectual abilities and a new theory about intellectual styles. While these contributions are inventive and likely to lead appropriately to a further broadening of our understandings of intelligence if used for research purposes, they are too avante guarde and speculative to be relied on in practice.

The Sternberg Multidimensional Abilities Test is based on the triarchic theory. It is necessarily, however, a limited approximation of the theoretically ideal assessment. The available empirical data in support of the test (p.236) are "favorable" but wholly inadequate if this test were being proposed as a substitute for the WISC-R in Special Education placements. Is the new test intended to select children for special programs? Or are classroom teachers supposed to assess and then train mental abilities? Sternberg is annoyingly vague about the purpose of the test except to say that it "will be suitable for students in Kindergarten through college, as well as adults," and that it is a group test to be administered in two class periods. Yet, test validity depends on the intended use. In later sections on assessment for special placements and for classroom instruction, I consider the type of evidence that would be required to support the use of a test for these purposes.

Intellectual style is proposed as an attribute independent of one's level of intelligence but which determines how intelligence is applied. For example, a scientist and a doctor might be equally bright, but one likes to solve novel problems and the other does not. Sternberg uses a governmental model to describe various styles. The scientist has a legislative style and likes to construct things and make rules, whereas the doctor has an executive style and likes to implement and follow rules. Similarly, styles can be distinguished as global or local, monarchic, hierarchic, oligarchic, or anarchic, and so forth. When all of the dimensions of the model are considered, one individual could be said to have a judicial, oligarchic, global, external, and conservative style of self-government. ("Bob" is at least correct when he portrays himself as one who likes to build systems.)

The governmental model of intellectual styles is an interesting metaphor. Although Sternberg posits style as if it were distinct from personality, the most useful aspect of the exposition is that it illustrates how intelligent behavior is mediated by personality. Whether this theory has something to offer apart from traditional theories of personality remains to be seen. This avenue of Sternberg's work does not have immediate implications for practice except to alert us to the danger that traditional assessments of ability might be biased toward one style, thereby underestimating the intelligence of individuals governed by other styles.

Lastly Sternberg considers the extent to which intelligence is socio-culturally defined. This is a key idea, essential to an understanding of how current scientific views have changed compared to historic theories of intelligence. The social construction of intelligence is considered further in the section on ability. Sternberg's explication of these issues has some troublesome features, however. First, he creates a false dichotomy between objective intelligence, which refers to the mental structures set by nature, and subjective, socially-defined perceptions of intelligence. Philosophers of science reject the positivist notion that there is such an objective, discernible truth independent of the scientist's subjective framework, constructs, and choice of instruments.

Earlier, Sternberg himself acknowledged that an attempt to assess intelligence from the perspective of the triarchic theory could not be made culture-free. In the main, Sternberg fails to draw the connections between his discussion of cultural definitions of intelligence in the last section and his earlier contextual subtheory; thus he allows for the influence of culture on the judgment of intelligence but not on its development.

CONSENSUS VIEWS: THE NEW PSYCHOLOGY OF INTELLECTUAL ABILITY AND LEARNING

In this section, consensus conclusions derived from the contributions in this volume and from cognitive psychology more generally are presented as summary statements followed by brief explanatory discussions. The stipulative summaries describe first the nature of intelligence and then the nature of learning and achievement according to contemporary psychological theory.

The Nature of Intelligence

Intelligence is not an inborn, permanent lump in each person's head. Intellectual ability can be developed to a great extent by opportunities to learn and think. The nature-nurture controversy has been prominent in public debates about intelligence for 20 years. Most educated individuals believe that intelligence is influenced by both heredity and environment. Nevertheless, lay conceptions of intelligence have not kept up with findings from developmental and cognitive psychology about the extent to which thinking ability is learned. Laymen tend to credit environment for the amount of school-relevant information an individual has been exposed to and for attitudinal differences caused by one child's family valuing education while another's does not. But lay views generally do not acknowledge that the mechanisms of intelligent thought are actually created as children are guided through their interactions with the environment by adults who model and explain things to them.

This is not to say that the authors in this volume dispute a genetic contribution to manifest intelligence. Rather, they deny that differences among individuals as presently observed are largely attributable to genetics and are therefore unalterable. The best example of the practical significance of this claim is offered by Brown et al.'s citation of Budoff's (1974) efforts to test the instructability of intelligence. As might be predicted by prior differences in opportunities to learn, Budoff found that among Educable Mentally Retarded subjects lower-class children were more likely to gain from instruction on concept-learning tasks than were middle-class children. Thus we make a mistake if we conclude that children who have not learned cannot learn.

In practice individuals who do not fully appreciate that children can learn to learn, are more willing than is warranted to assign individuals to different

opportunity tracks, like special education classes for the mildly handicapped or average instead of gifted programs, on the basis of what Brown et al. call "static" measures of ability. We have also found that school board members and educators may be willing to act on nativist beliefs regarding cognitive abilities when the policy decision appears in a guise different from IQ testing (Shepard & Smith, 1985; Smith & Shepard, 1988). For example, an estimated 20% of school districts nationally use some kind of "readiness" measure to determine which children may enter kindergarten or first grade. An environmental perspective would suggest that children who score poorly should be in school to have access to learning opportunities that will develop language and learning concepts, whereas the nativist position supporting these policies argues that biological readiness cannot be hurried and therefore low scoring children are better off waiting a year to allow these abilities to emerge spontaneously. Because the tests associated with readiness policies are not called ability tests, decision makers may not even realize that what they have learned about the heritability of "intelligence" is pertinent.

Intelligent thought involves "metacognition" or self monitoring of learning and thinking processes. If policy makers wanted to learn one piece of scientific jargon that best summarizes the contribution of cognitive psychology to today's understanding of intelligence, they would learn the term "metacognition." Whereas earlier generations of psychologists had treated internal brain processes as if they were unknowable black-box mechanisms, the goal of cognitive science has been to examine and model these internal processes. What they have learned is that intelligence is a set of hierarchically organized mental activities that enable the individual not only to solve problems but to monitor and direct problem solving.

Metacognition is the general term referring to the next level of thinking, or thinking about thinking, which includes a variety of self-awareness processes. Intelligent though involves these higher-order processes identified by Brown et al. as "the ability to allocate one's mental efforts efficiently, to plan, monitor, oversee, orchestrate and control one's own learning." Sternberg is talking about the same sorts of things with the metacomponents and knowledge-acquisition components in his internal subtheory. Sternberg called his metacomponents of intelligence "executive processes" which include: (a) recognizing the existence of a problem, (b) deciding upon the nature of the problem, (c) selecting a set of lower-order processes to solve the problem, (d) selecting a strategy to combine these components, (e) selecting a mental representation of the problem, (f) allocating one's mental resources, (g)monitoring one's problem solving as it is happening, and (h) evaluating problem solving after it is done.

In practical terms, metacognition is important because it is the development of metacognitive abilities, rather than lower-order skills, that is more likely to make an individual more intelligent. When cognitive psychologists say that intelligence is developed, they mean most significantly

that these learning-to-learn processes are acquired through experience. Furthermore, when an individual fails to learn these abilities "naturally" through their own powers of reflection, they can be instructed explicitly. For example, Brown has been able to teach children how to comprehend better and Sternberg has trained adults how to learn vocabulary in natural-language contexts.

Intelligence is socially and culturally constructed. The more scientists learn about the mental mechanism we call intelligence, the less sense it makes to think of them as content-free processes. Intelligence is socially developed in the same way that language, gestures, interpersonal behaviors, manners, tastes, etc., are shaped by a child's interactions with family and community. According to Sternberg, the mental process that constitute intelligence may be the same across cultures but their instantiations will be radically different in different contexts. Culturally determined conceptions of intelligence not only govern want is valued in a society but actually shape the development of different mental abilities, as in the example cited by Sternberg where aboriginal children have more highly developed spatial memory strategies than do Anglo-Australian children who rely on verbal strategies for the same memory tasks.

This insight from cognitive psychology--that metacognitive processes are developed by one's culture in the same way that language structures and concepts are developed--has profound implications not only for the assessment of intelligence but for the provision of learning opportunities to children with different cultural experiences. This research changes the meaning of terms like "culturally deprived" and "culturally disadvantaged." Feuerstein's work, cited in the Brown et al. chapter, allows for the possibility that children could be culturally deprived if their interactions with adults are so limited that they are never socialized into their own culture. Except for extreme cases however of parent absence or apathy, it is not true to say that poor and minority children in the United States do not have a culture or that they have not developed intelligent thought processes according to the constructs of their culture. These children are not culturally deprived or disadvantaged, they are culturally different. The insights offered by cognitive psychology jibe with those of numerous sociological studies which suggest that children of diverse cultural backgrounds do poorly in American schools not because they have no culture but because of the mismatch between their frame of reference and that of the dominant culture. If teachers could engage students in learning in ways more compatible with their own cultural patterns, fewer students would appear to be unable to learn.

Although Sternberg discusses Shirley Brice Heath's work extensively to illustrate the influence of community modes of discourse and learning on the development of thought processes, even he does not see the full implications of Heath's findings. In recounting Heath's descriptions of child rearing practices in a black working-class community she called Trackton, Sternberg emphasizes the lack of academic preparation of these children rather than their competence within

their own culture. The tone of Sternberg's rendition preserves the idea that children from lower socio-economic circumstances come from deficient cultures. He does not suggest how they can use their community-specific intelligence to learn in school or even whether he thinks this is possible. A more careful reading of Heath's original work (Heath, 1982), reveals that many of the school language expectations that put poor black children at a disadvantage are arbitrary, determined by the pattern of discourse familiar in middle-class homes, and not fundamental to legitimate academic goals. For example, Heath documents how pervasive the use of questions is in classroom language routines. Often white middle-class teachers address questions to children even when they mean to make a declarative statement. They also require children to demonstrate comprehension by asking them to name discrete objects removed from their natural context whereas Trackton children would be more comfortable retelling about things in a story context. Ironically requiring children to enumerate things, rather than relating meaning and connecting meaning to their own experiences, is not consistent with sound learning theory for any children and needlessly places culturally different children at a disadvantage. When Heath worked with teachers to try to adapt initial instruction to language forms more familiar to Trackton children, they found they could elicit far more energetic talk and lesson participation by asking higher-order "probing question" than by asking for naming of objects. They also found that they could teach children explicitly about the kinds of questions and answers that were expected in school rather than presuming that the children were deficient if they did not come to school already knowing these conventions.

Intelligence is multifaceted. Individuals have not one but several intelligences of different degrees of strength and development. Although psychologist would not all agree on the specific subtypes of intellectual ability, there is much wider agreement that intelligence is indeed differentiated. An individual can be very "smart" in one type of endeavor and quiet dull in another. Even individuals who are regarded as generally very bright show great differences among their own strengths and weaknesses and retarded individuals, except for those who are profoundly disabled, have profiles showing varying degrees of competence in different areas. Again, intelligence is not an undifferentiated lump of a certain size that governs the amount of a person's ability for all time and all tasks.

Evidence to support the multidimensional nature of intelligence comes from several lines of inquiry including the componential analyses exemplified by Sternberg's earlier work. Gardner based his conclusions about the existence of multiple intelligences on the fact that normal children are not equally accomplished in all areas and the finding that brain-damaged individuals may lose function in specific areas without harm to their other abilities. Cross-cultural research, demonstrating that different mental abilities are developed to a greater or lesser degree in different cultures, also supports the idea that individuals have

different kinds of intelligence not just different amounts of "it". In fact, the argument that intelligence is multifaceted overlaps with the previous point that intelligence is socially constructed, except that additionally individuals within a culture will nonetheless show different patterns of strengths and weaknesses. Although Gardner casts a very broad net, including bodily-kinesthetic thinking and interpersonal knowledge as types of intelligence, the idea of multiple dimensions of thinking ability has important educational consequences even if one focuses only on the more traditional academic abilities of linguistic intelligence, logical-mathematical ability, and spatial reasoning. To the extent that individuals have different patterns of strengths and weaknesses in these areas, they will be helped or hurt by modes of instruction that presume only one approach or one pathway to understanding. For example, statistics is a quantitative subject that draws heavily on one's logical-mathematical ability. However, I have found that graduate students in education and psychology, who tend to be verbal learners, can develop sophisticated conceptual understanding of statistical methods if relations are explained verbally rather than expecting that they are "obvious" from the equations.

Also, in practical terms, the existence of multiple intelligences means that individuals cannot be reliability ranked on a single continuum. Therefore, school placement practices that separate children for differently paced instruction according to one dimension of ability (usually verbal-reading ability) will clearly misassign children according to their other abilities.

The Nature of Learning and Achievement

There is not a neat and tidy distinction between developing intelligence and learning to think about subject matter. To organize this commentary, I have adopted the traditional distinction between intelligence and achievement because these are the terms familiar to laymen. (If one of the learning principles offered in this section is that new learning ought to build on students' background knowledge, then it seems advisable to start with what is known.) A careful reading of these chapters, however, should make it difficult to maintain the separation. The development of metacognitive or self-monitoring processes resembles very closely the critical thinking processes that the Resnicks discuss in the context of subject matter expertise. Is the ability to analyze and comprehend what one reads a sign of intelligence or reading achievement?

Of course, a distinction can still be made at the extremes between general cognitive strategies that could be applied in several learning contexts and domain specific knowledge. But there is considerable overlap between the constructs of intelligence and achievement, when one is talking about conceptual understanding and ability to apply knowledge of subject matter. One pragmatic consequence of the blurred distinction between developed learning ability and higher-order learning of subject matter, is the conclusion arrived at by Brown et

al., the Resnicks, and implicitly by Gardner: the development of critical thinking can be accomplished in the context of important substantive topics. There is no reason to digress and practice thinking about puzzles that bear no relation to real experiences. Note that many of Sternberg's ability tasks run contrary to this principle. For purposes of measuring one's ability to cope with novelty it might make sense to pose problems based on counterfactuals. It does not follow, however, that instruction based on these types of tasks would have the desired amount of transfer, i.e., would help the individual become better at solving anything other than this particular type of problem.

By-rote sequential instruction does not foster critical thinking or meaningful learning. To be perfectly clear about how revolutionary and important the findings from current psychological research are to the future of education, it is useful to explain what the "outmoded learning theory" is, how it is carried forward in numerous public policies, and why it does harm to student learning. The Resnicks provide a succinct summary of what scientists believed about learning 50 years ago. The decomposability assumption refers to the idea that learning can be taken apart into its constituent bits and transmitted to the new learner bit by bit. The real spirit of the old theory, sometimes referred to as behaviorist or associationist theory, is captured best by a couple of popular analogies. According to the old view, sometimes called the "bean-jar theory of learning," the student is an empty vessel into which knowledge is poured, one bean at a time. Others prefer the tower of blocks metaphor--learning occurs by stacking blocks of information one on top of the other--because it also implies that there is a prescribed sequence for acquiring bits of knowledge. The old theory made no provision for how insight or conceptual understanding was to take place, nor did it admit that there needed to be any organization in the mind to comprehend or use information.

As illustrated graphically by the Resnicks and Gardner, the old learning theory is enforced in present day practice by standardized tests. If tests administered for accountability purposes exert pressure on classroom teachers, then implicitly day-to-day instruction is governed by the old learning theory underlying tests even if teachers nominally have a more up-to-date understanding of how children learn. Specifically, when teachers drill on facts, use worksheets resembling standardized tests, substitute multiple-choice items over essay tests for their own classroom evaluation, and reduce the amount of time spent on activity-based learning and problem solving, the old learning theory can be said to have driven out good instruction.

In addition, educational policies that determine what should be done when students are doing poorly in school almost invariably enforce implementation of the old theory. For example, the back-to-basics movement not only stresses the importance of basic academic skills but assumes a sequential bit-by-bit learning model. If students have not mastered the basics, such as number facts in mathematics, they cannot go on to conceptual problems. From the vantage point

of current research discussed in the next section, we see that children who are behind are then doubly disadvantaged because they are denied any kind of context that would make the number skills more meaningful and easier to learn. Grade-to-grade promotion tests, grade retention, and skill-drill remediation programs are all educational policies premised implicitly on the outmoded theory of learning. Ironically, the sequential, bit-by-bit learning model does the most harm to slow learners because it postpones sometimes indefinitely opportunities for these children to learn things that are intrinsically interesting and connected to their own real-life experiences.

Learning is a constructive process. The learner must build a schema in her own head to understand a body of knowledge. Many people have had the experience of failing to learn the directions to a distant destination if each time they visit it they travel as a passenger; but they learn the directions well as soon as they drive to the destination themselves. Learning is an active process. Rather than passively receiving information (without noting landmarks or changes in direction), the learner must actively make sense of new knowledge and decide how to integrate it with previously held concepts and information. The learner must make meaning. Many metaphors have been used to describe the constructive nature of learning, including the idea that mastery occurs as an individual develops their own conceptual map of a knowledge domain.

The Resnicks make the important point that students must think and interpret even to learn simple reading and math skills, otherwise they perceive only disembodied and nonsensical strings of words and numbers. To comprehend a reading passage, the reader develops a mental image or outline of key points. Good readers ask questions of the text and reread when they don't understand. Children learning mathematics must invent mental models to represent arithmetic operations. If, by dint of extraordinary effort, children only memorize rules without understanding, the "knowledge" is of no use because it can't be retrieved, applied, or generalized. Meaning makes learning easier, because the learner knows where to put things in her mental framework, and meaning makes knowledge useful because likely purposes and applications are already part of the understanding.

Effective instruction helps students to use what they already know to arrive at new understandings. If real learning requires the student to make sense of things, then teachers telling the answers will often fail to produce student learning. Instead, good teachers stage activities that allow students to make connections and gain insight for themselves. This does not mean that students are allowed to wander aimlessly. Instructional opportunities must be carefully guided, to use Brown's term, i.e., tailored to the student's level of readiness and monitored to ensure that learning occurs. Instruction designed to make students do the mental work is often referred to as scaffolded instruction. The teacher provides the external support--not unlike an adult holding out two

fingers to support an infant's first steps--until the student can perform the conceptual task independently.

Effective instruction also focuses on making meaning by engaging students in purposeful tasks, eliciting student background knowledge, and teaching students explicitly about learning strategies. If a goal of teaching mathematics is for students to be competent at everyday tasks, like making change or buying the right amount of paint to apply two coats to a 10' by 12' room, then these are the types of experiences that young children should have as the context for learning addition and subtraction facts, the decimal system, and multiplication. Their intuitions about how to solve simple problems should be the basis for eliciting and discussing their mental models; the goal of teaching is to help students extend their conceptual understanding rather than telling them the rules. By explicitly teaching metacognitive activities such as Brown et al.'s strategies of questioning, summarizing, clarifying, and predicting, students receive permission not to know "the answer" immediately and learn that there are intermediate steps to understanding. Without this explicit modeling students who don't "get it" easily are apt to gaze at the ceiling hoping that a ready-made answer will pop into their heads. In essence, if you want students to be able to think when they're finished, teach them to think every day in each school subject.

IMPLICATIONS FOR ASSESSMENT

What's wrong with traditional tests?

All of the authors in this volume have explained the important features of their own work in contrast to traditional test theory, which has been harmful because of its oversimplified and distorted assumptions about human learning. Traditional efforts to measure intelligence have erred by assuming that learning potential could be assessed by static, one-dimensional, culturally neutral instruments. Traditional achievement tests have suffered from the behaviorist notion that learning can be carved up into discrete instructional objectives and from psychometric reliance on a correlational model. According to the latter theory, tests are considered to be adequate measures of a knowledge domain if they are highly correlated with more in-depth assessments of that domain. This is the (concurrent validity) argument used, for example, to substitute short tests for long tests and, many years ago, to substitute multiple-choice tests for essay tests (see Coffman, 1971). Now we see in a different political context that the adequacy of multiple-choice tests as proxies for more in-depth assessments of student learning is seriously in doubt--both because we distrust the accuracy of the data (do the scores really tell us what students know?) and because of the influence of politically important tests in redirecting what is taught.

Policy makers are undoubtedly tired of hearing complaints about standardized tests. Unless one is persuaded about the seriousness of harm that can be done by ill-conceived tests, the criticisms sound like so much whining from educators who don't want to be held accountable. The incessant repetition of what's wrong with testing is useful in two respects: 1) although there is some movement to adopt assessment programs more conducive to student learning, most policies and practices have not changed--suggesting that the insights offered in this volume are not yet widely understood; and 2) measurement specialists and cognitive psychologists alike know better what is wrong with the old tests than precisely how to make the new assessments. Thus the complaints help to clarify the principles that should guide the reform of assessment practices. What is sketched in the following sections, in broad outline, is the character to be aimed for if assessments are shaped by current perspectives on learning potential and learning progress.

Assessments tailored to educational purposes

To translate important discoveries from cognitive psychology into recommendations for assessment requires an understanding of purpose. The Resnicks explain how the features of testing programs must necessarily vary with function. They identify several dimensions of variation: the audience for test results; the extent to which information is needed for individual students; the need for independence from teacher's judgments; how quickly the data must be available; and the level of detail required in the test results. For example, teacher observations of reading performance are the most detailed and most timely way to inform day-to-day instructional decisions but are not trusted as objective evidence for accountability purposes.

Similarly it has also always been the case that psychometric criteria for judging the adequacy of measuring instruments depend on purpose. Two dimensions govern the stringency of technical requirements for reliability and validity evidence. First, is the test to be used to make individual or group decisions? Group data can tolerate less precise measurement because, statistically speaking, group means are stable even when individual scores are not. Second, will test results be used to make crucial decisions, like placing a student in Special Education or a school district in receivership, or not so important decisions, like counseling a student about career options or planning curriculum improvement for the next year? Tests used to make irreversible, important decisions about individual pupils require the greatest degree of technical accuracy. In addition, the type of validity evidence sought depends on the particular use made of the test results. For example, an early childhood measure involving language concepts and gross motor skills might be very useful to kindergarten teachers for planning instruction but invalid for screening children into transitional first grade classrooms because the test is not accurate enough to categorize children as "ready" and "unready" and because the placement itself (the

transition grade) does not have validity evidence of effectiveness (Graue & Shepard, 1989). Implications from cognitive learning theory for assessment are therefore considered separately for each of several educational purposes.

Implications for Classroom Assessment

Gardner and Brown et al. talk about assessment intended to assist the individual learner. If we have an image of the teacher guiding the student just as a parent guides a child, to try new things just within his reach of understanding, the teacher has need of constant information both about what the student knows and the strategies being used to process and comprehend new concepts. For purposes of instruction, the ideal assessment instrument is the mind of the teacher so that decisions and correctives can occur "on the fly" as Garner put it. By imbedding diagnostic assessments in instructional activities, teachers can preserve the integrity of assessment tasks (the wholeness of tasks and natural learning context) and protect instructional time that would otherwise be diverted to testing.

Although it is very clear that informal day-to-day assessment is essential to effective teaching, it is also painfully clear that teachers presently do not have the training to carry out assessments of mental processes with the kinds of insights envisioned in this volume. Rather, teachers have been trained in a "follow the book," test-teach-test approach (Calfee & Hiebert, 1988) which identifies missing facts but not specifics about failures in a student's thinking. Nonetheless, it is my position here that there are no viable alternatives except to train teachers better.

There is general agreement that external, packaged tests will not solve the problems of what teachers need to know about student learning. Most external tests such as state-administered criterion-referenced tests or commercially published diagnostic batteries are simply too cumbersome, as well as sharing the conceptual limitations of other multiple-choice tests. Important instructional decisions are made every day. It is simply impossible for formal, uniform tests (where every student in a grade takes the same test) to inform these decisions unless teachers spend as much time testing as teaching. The divisions of assessment purposes made by the Resnicks imply critically important policy choices. Decision makers cannot imagine that they can mandate a single test to serve both accountability and instructional purposes. Any test tells only what students know on the day of the test, and once the data are removed from the context of observation, it tells very little even about how they know what they know. Therefore, accountability tests would be minimally useful to instruction only on the day after testing. Furthermore, if tests are broad enough to cover the full range of what students know, they cannot possibly be detailed enough to be diagnostic about the comprehension of individual students. When the Resnicks discuss the instructional importance of thinking tasks in accountability assessments they are not interested in the data that will be reported back about

individual pupils, they are interested in the symbolic importance of assessment tasks in shaping what teachers throughout the system chose to teach their students.

Finally, an important implication for classroom assessments pertains to the choice between teaching thinking in the context of subject matter or teaching thinking with novel cognitive tasks. In the past, Special Education in particular has had some bad experiences when models of internal brain processes were turned into instructional programs intended to fix the brain. Years later research summaries confirmed that training on underlying process tasks improved scores on those tasks but did not transfer improvement to target academic tasks (Arter & Jenkins, 1979). As a result time spent in Special Education training actually diverted time from academic work rather than improving it. Feuerstein specifically used non-school tasks because he wished to avoid topics for which deficient students had already developed a considerable aversion. Obviously, then, there are arguments on both sides. Given the risk that "training abilities" without transfer can divert attention from school topics that are relevant and useful, the burden of proof should rest with those who want to train students with exotic tasks that are unfamiliar and bear no relevance to everyday problem solving. When Brown et al. propose to work with very young students on academic tasks they have the best of both worlds, i.e., to intervene and teach thinking in the context of school subjects before students have had negative experiences with school content.

Implications for Placement and Tracking

The traditional view of intellectual ability as a generalized, fixed trait led to a convenient model for assigning individuals to their permanent place in the educational system. Individuals could be tested and located on a continuum describing their ability to learn and hence the pace at which they should receive standard instruction. Those at the very lowest end of the scale were placed in Special Education, and the next-lowest into low-ability tracks. Those at the top end of the scale were placed in high-ability tracks or gifted and talented programs. This model has never admitted the possibility that instruction could be aimed at developing the capacity to learn nor that an individual's location on the continuum might be inaccurately assessed. Despite three decades of research on the negative effects of tracking, this model still dominates much of educational practice (Oakes, 1985; Slavin, 1987).

On the face of it, the authors in this volume do not address themselves to problems of Special Education diagnosis, or assessment for purposes of remedial instruction or class placement. In fact, however, the evidence they have amassed speaks directly to the matter--leading unequivocally to a rejection of the fixed-ability special-assignment model. First, it should be clear from the extensive research recapitulated in the chapters that traditional, static measurements of

learning ability are inaccurate for individuals who have had inadequate opportunities to learn. Second, the instructional treatments given to individuals in Special Education and low ability tracks can be expected to be ineffective because they are based on an outmoded learning theories--that emphasize by-rote learning of skills and that intentionally postpone opportunities to learn to think until after basic skills have been mastered.

Educational researchers and sociologists have for years documented the negative effects of tracking, Special Education placement for mildly handicapped children (Carlberg & Kavale, 1980; Madden & Slavin, 1983), and grade retention (Shepard & Smith, 1989). These separate research literatures have several key features in common. Although each educational "treatment" is intended to improve academic achievement by putting at-risk students in a place where instruction will be aimed more closely to their learning level, controlled studies show that members of each poor achieving group learn more when placed with regular students than when they are separately placed and get a special treatment. Research also consistently finds that each type of special placement carries a negative social stigma and, in each case, that the intended individualized instruction is not accomplished by the special placement.

Insights from cognitive psychology are able to provide a much more concrete explanation as to why at-risk students would consistently learn less from the kinds of basic-skills remediation they typically receive. In the past, sociologists have documented that children in Special Education and low ability tracks are given a "watered-down" curriculum. However, because the public shares with most educators the linear-learning idea that students can't learn "hard" things until they know the basics, policy makers have not been able to see concretely what instructional alternatives exist to the go-slow model. The Resnicks, Gardner, and especially Brown et al. in this volume provide a rich description of the kind of instruction students should receive if we want to improve student learning dramatically. Ironically, in today's schools only gifted children regularly enjoy opportunities to develop thinking skills by working on contextualized problems and extended projects, by having critical thinking and questioning modeled for them, and by trying out and practicing their own reasoning efforts in social settings.

The authors in this volume do not tell how to make special placement assessments more accurately. Instead they offer a view of learning and effective instruction that recommend against special placements for all but the most severely handicapped students.

Policy makers endorse and perpetuate a fixed-ability tracking model of education when they make decisions to require nonpromotion of students who are deficient in basic skills, exclude children from kindergarten or first grade who are "unready to learn," require that at-risk or remedial students be in self-contained programs to receive funding, and forbid the co-mingling of Special Education funds to support instruction of handicapped children in the regular classroom.

Therefore, the understandings of learning offered in this volume would have a profound effect if they could alter the fixed-ability special-assignment view of education that is still so powerful in the public mind.

Implications for Accountability Assessment

Following the advice of the Resnicks, state legislators and school board members should be mindful of the effects mandated accountability tests will have on what is taught in schools. In the current political climate of intense attention to test scores, standardized multiple-choice tests have become templates for instruction. Whatever can be said about the adequacy or efficiency of multiple-choice tests as indicators of student achievement, multiple-choice tests cannot be defended as good curriculum. "Teaching to the test" as practiced by most educators is not the same as cheating, which would involve practicing on the exact questions from the test or giving students answers to test questions. "Teaching to the test" means letting the content of the test and even the types of questions on the test become the exclusive focus of instruction. As a consequence teachers stop giving essay tests or having students work on projects and instead spend the year on fill-in-the-blank worksheets. Although accountability pressure is intended to improve the quality of education, it may actually worsen student achievement by driving out opportunities that develop thinking and reasoning abilities.

If accountability measures determine what is taught and we want students to learn to think, then assessment exercises must be devised to be much more ambitious and extended tasks requiring students to demonstrate the kinds of reasoning and problem solving that are the real goals of education. The point is not just to ensure more valid measurement but to redirect instruction toward more challenging learning goals. The Resnicks describe several types of performance assessments as alternatives to traditional tests. The key is to capture in the assessment tasks demonstration of the specific performances, reasoning abilities, oral and written communication skills, etc., that we wish students to acquire as outcomes of education. Various efforts to reform assessment use terms such as "authentic" (Wiggins, 1989), "direct" (Frederiksen & Collins, 1989), and "performance" assessment to convey the idea that assessments must capture real learning activities if they are to avoid distorting instruction. Proxy measures like standardized tests invite distortion because it is possible to raise test scores without genuinely improving student achievement (Shepard, 1989).

Performance measures are much more difficult to develop and score than conventional tests. Nonetheless the examples reviewed by the Resnicks suggest that such measures can be scored reliably enough to provide accurate accountability evidence. Generally authentic or performance assessments use one of two strategies. Either tasks are constructed, like writing prompts or science experiments, that can be administered under standardized conditions, or work

products are collected from the students' on-going instruction, such as portfolios of writing assignments or math papers. To permit aggregation and comparison of results scoring rules must be developed for rating written products or the quality of students' responses when interviewed after a science experiment. When portfolios are used, safeguards such as occasional audits are required to ensure that students submit their own work and that instruction is not distorted by spending all year polishing a single assignment.

Scoring by human judges rather than optical scanning machines costs more. The reason policy makers should be willing to invest in performance assessment is, again, not just because it will yield more valid data but because with the right kinds of tasks, it will lead educational reform in the right direction. Given the current negative effects of testing on instruction, the urgency for reform of assessment is great. Performance assessments could be undertaken without increasing the budgets currently spent on state and local testing if policy makers were willing to trade off and test fewer students, using scientific sampling procedures, or fewer grades and subject areas (Resnicks, this volume; Shepard, 1989). The idea that accountability requires testing every pupil in every grade in every subject has to be given up to make it feasible to institute performance assessments.

It should be clear for logistical reasons that the same assessment cannot be used for instructional purposes and for accountability to external audiences. The views offered by Brown et al. and the Resnicks as to how assessment should be undertaken at these two levels are compatible but not the same. They are informed by the same learning theory and would engage students in the same kinds of activities aimed at developing their ability to think about subject matter. But in classrooms the teacher should be the one who assesses and guides instruction. Conversely for accountability purposes, teachers can be trained to score assessment responses but should not be responsible for judging their own students. A collateral benefit of large-scale performance assessments, in fact, is the amount of professional development that occurs during training and scoring when teachers see student efforts from other classrooms and are asked to reflect on their own goals and criteria for judging the quality of student work.

CONCLUSION

Since the time that most of today's policy makers were in school, our research-based understandings of human intelligence and learning have changed profoundly. Two of the most important findings from cognitive psychology are: 1) that intelligence is developed, and 2) that all learning requires thinking. Furthermore, activities and socially-supported interactions that develop intelligence are virtually indistinguishable from the kinds of instruction that enable students to think critically about subject matter.

Unfortunately much of educational practice, most especially traditional standardized tests, carry forward the assumptions of psychological theories that have since been disproven. Students are still placed in special educational programs, low-ability tracks, readiness rooms, and the like on the basis of static measurements that do not take account of opportunity to learn in calculating ability to profit from instruction. A linear, bit-by-bit skills model of learning is enforced so that students are not allowed to go on to thinking until they have mastered the basics. In the name of accountability and educational reform teachers are forced to spend so much time on the multiple-choice format curriculum that they provide students few of the opportunities that would teach them to think.

What the insights in this volume offer is a reconceptualization of instruction to teach thinking and a commensurate reformulation of assessment to help not hinder that effort.

REFERENCES

Arter, T.A., and J.R. Jenkins. 1979. Differential diagnosis--prescriptive teaching: A critical appraisal. Review of Educational Research, 49:517–555.

Budoff, M. 1974. Learning potential and educability among the educable mentally retarded. Final Report Project No. 312312. Cambridge, MA: Research Institute for Educational Problems, Cambridge Mental Health Association.

Calfee, R., and E. Hiebert. 1988. The teacher's role in using assessment to improve learning. In Assessment in the service of learning: Proceedings of the 1987 ETS Invitational Conference. Princeton, NJ: Educational Testing Service.

Carlberg, C., and K. Kavale. 1980. The efficacy of special versus regular class placement for exceptional children: A meta-analysis. Journal of Special Education, 14:295–309.

Coffman, W.E. 1971. Essay examinations. In R.L. Thorndike (Ed.), Educational Measurement, Second Edition. Washington, DC: American Council on Education.

Frederiksen, J.R., and A. Collins. 1989. A systems approach to educational testing. Educational Researcher, 18:27–32.

Graue, M.E., and L.A. Shepard, 1989. Predictive validity of the Gesell School Readiness Tests. Early Childhood Research Quarterly, 4:303–315.

Heath, S.B. 1982. Questioning at home and at school: A comparative study. In G. Spindler (Ed.), Doing the ethnography of schooling: Educational anthropology in action. New York: Holt, Rinehart and Winston. Madden, N.A., and Slavin, R.E. (1983). Mainstreaming students with mild handicaps: Academic and social outcomes. Review of Educational Research, 53:519–569.

National Council of Teachers of Mathematics. 1989. Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Research Council. 1989. *Everybody counts--A report to the nation on the future of mathematics education.* Washington, D.C.: National Academy Press.

Oakes, J. 1985. Keeping track. New Haven, CT: Yale University Press.

Palincsar, A.S., and Brown, A.L. 1984. Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. Cognition and Instruction, 1(2):117–175.

Shepard, L. 1983. The role of measurement in educational policy: Lessons from the identification of learning disabilities. Educational Measurement: Issues and Practice, 2:4–8.
Shepard, L.A. 1989. Why we need better assessments. Educational Leadership, 46:4–9.
Shepard, L.A., and M.L. Smith. March 1985. Boulder Valley kindergarten study: Retention practices and retention effects. Boulder, CO: Boulder Valley Public Schools.
Shepard, L.A., and M.L. Smith. 1989. Flunking Grades: Research and Policies on Retention. New York: The Falmer Press.
Slavin, R.E. 1987. Ability grouping and student achievement in elementary schools: A best-evidence synthesis, Review of Educational Research, 57:293–336.
Smith, M.L., and L.A. Shepard. 1988. Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. American Educational Research Journal, 25:307–333.
Wiggins, G. 1989. Teaching to the (authentic) test. Educational Leadership, 46:41–47.

INDEX