Beliefs about differences between ability and accomplishment: From folk theories to cognitive science

David F. Lohman[‡]
University of Iowa

Beliefs about the extent to which intelligence can be disentangled from achievement pervade both theorizing about the nature of academic giftedness and efforts to identify those children most likely to display it. I review some of the evidence that challenges common assumptions about the independence of these constructs. As in many other domains, experts hold views that differ markedly from those of novices. Most novices believe that ability is innate and achievement acquired, whereas experts see the two as different aspects of the same thing. A better understanding of the unique and common aspects of measures of achievement and ability can improve both how students are identified for inclusion in programs for the academically gifted and the objectives of such programs.

I sat there squirming as he talked. The speaker, a well-traveled professor from another university, criticized what he termed "so-called ability tests" that presented tasks that required knowledge of language or any other learned symbol systems. The vocabulary subtest of the Stanford-Binet (Roid, 2003) was first in line for censure. "How could a smart child who does not speak English possibly know the answer to these questions?" he asked. The Verbal Comprehension subtest on the Wechsler Intelligence Scale for Children (WISC; Wechsler, 2003) was next. An item from the test appeared on the screen. Then, a similar phrase from a Boy Scout manual appeared next to it. "Do you get a better WISC IQ score if you study your Boy Scout manual?" he asked. The group-administered ability test on which I work was next in line. "How could one possibly estimate the ability of a child by asking her to read and complete sentences with missing words, or make judgments about the relative size of 2 + 3 and 2 × 3? Are not these learned skills?" However, his most scathing criticism was reserved for an item on the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001). The same vocabulary item was used both in the Verbal Comprehension test on the ability battery and in the Reading Vocabulary test on the achievement battery. "How can the same item measure both ability and achievement?" he shouted. A good ability test, he said, should contain only those items that could be administered with equal fairness to an individual who had no access to language or school as to the individual immersed in the dominant culture. How is this possible? His solution was to measure ability using only nonverbal, figural reasoning tests like the one on the handout he had distributed. Most of the teachers of gifted children in the audience seemed to find his arguments congruent with their beliefs about the matter. A few squirmed in their seats, but, like me, did not raise any of the questions that begged to be answered. Perhaps at some level they too were troubled by this skillful use of verbal persuasion to deny the importance of language.

The speaker's arguments struck a familiar chord because most people believe that that intellectual capacity can be separated from intellectual accomplishments. Indeed, this belief is fundamental to many conceptions of giftedness. For example, Gagné (2003) distinguished between *gifts* (i.e.," untrained… natural abilities," p. 60) and *talents* (i.e., "developed abilities [or skills] and knowledge," p. 60). Tannenbaum (2003) distinguished between the promise of childhood abilities and productive adult giftedness evidenced by superior accomplishments. In her survey of procedures commonly used for identifying gifted children, Assouline (2003) noted that many school psychologists recommend using an ability test to identify gifted children and an achievement test to guide educational programming. The common theme, then, is that one can distinguish potential for accomplishment from accomplishment.

However, potential is not always seen as easy to measure. Gardner (2003) has eschewed the difficult task of measuring the potentials he has described. Others, who value the sort of cognitive potential measured by intelligence tests, are also quick to point out the limitations of such measures (e.g., Tannenbaum, 2003). At least for these theorists, it is

[‡] David Lohman, 366 Lindquist Center, The University of Iowa, Iowa City IA 52242. Email: david-lohman@uiowa.edu Fax: (319) 335-6145.

David Lohman is Professor of Educational Psychology at the University of Iowa. His interests are the nature, measurement, and educational implications of human cognitive abilities. He is the former chair of the Division of Psychological & Quantitative Foundations in the College of Education, and with Elizabeth Hagen, co-authors the Cognitive Abilities Test. Copies of recent papers may be obtained at http://faculty.education.uiowa.edu.

easier to make the distinction between potential and accomplishment than to measure it.

Can we have our potential and measure it too? More concretely, how should we understand the difference between these constructs? Can we measure both? Neither? Although the distinction between potential and accomplishment can be made in any domain in which one can define expertise, it is most commonly invoked in discussions of academic giftedness. It is here that I will focus my attention. My themes are that (a) commonly-held beliefs about the separability of ability and achievement are much closer to folk than to scientific theories of these constructs, (b) the development of beliefs about the relationships between ability and achievement follow a predictable sequence as individuals gain more understanding of each, and (c) better understanding of the enmeshment of these constructs can improve the both the process by which gifted students are identified and the goals of the programs designed to serve them.

The Jangle Fallacy

In a book published in 1927, Truman Kelley called attention to something he dubbed the "jangle fallacy" (p. 63). Kelley was the lead author of the first edition of the Stanford Achievement Test (Kelley, Ruch, & Terman, 1922). As a statistician of some repute, Kelley was bothered by the way people treated scores on his achievement test and various intelligence tests as if they measured independent constructs. He knew that the overlap in individual differences on the two types of tests was enormous. For the tests he was using, Kelley estimated that about 90 percent of the "true" or systematic variation in general intelligence was shared with composite measures of academic achievement (see Lubinski, 2004, for an explanation of why composite scores generally show the most overlap).

The culprit, he said, was language. Because different words—*intelligence* and *achievement*—were used to describe the constructs measured by the two types of tests, people treated intelligence and achievement scores as if they were in fact distinct. "The glibness with which we differentiate between achievement and intelligence is explained in part by the fact that our language is at fault" (Kelley, 1927, p. 63). He then referred to something E. L. Thorndike (1903) had called the *jingle* fallacy. This occurred when people treated as identical members of a group that merely shared a similar label. Kelley coined the phrase *jangle fallacy* to describe the opposite problem, that is, the tendency to treat terms that sound different as if they really signified different concepts. *Intelligence* and *achievement* sound as though they are (or should be) different things. Therefore we treat measures so labeled as if they measured different things.

Kelley was surely at least half right. Language indeed abets our propensity to view concepts that differ only in degree as differing in kind (Nickerson, 2004). But language not only shapes thought, it is in turn shaped by the concepts and belief systems we have constructed and wish to express. The larger problem, then, may be the beliefs themselves.

A Theory of Personal Theories of Ability and Achievement

Like most fallacies in educational measurement, the jangle fallacy seems to recur in every generation. In large measure, this is because those of us who study abilities and achievements go through a similar progression in our beliefs about the nature and nurture of intellectual competence. These are not stages. Learners construct increasingly sophisticated theories in all domains of nontrivial complexity as their expertise grows (Duit, 1991). As in most such developmental schemes, virtually everyone begins with a version of the simplest theory. Why some move on while others remain committed to a particular belief is not always clear. One factor that seems to matter is a willingness to consider (and even to seek out) evidence that contradicts one's current views. Openness to new perspectives is difficult if one has a vested interest in preserving the current belief. Cognitive styles or even a more general growth in epistemic cognition (Hofer & Pintrich, 2002) may also matter. A fondness for the sort of sharp categories that typify simpler theories may make it difficult to move to fuzzier worlds in which there is more grey than black and white.

There are several clear examples of this progression in beliefs about the constructs of intelligence and achievement in the history of educational measurement. In his earliest writings, E. L. Thorndike (1903) spoke optimistically of the possibility of separating the two. However, after another 20 years of research, "Thorndike saw the problem of defining and measuring intelligence as increasingly complex and the solution receding farther into the distance" (R. M. Thorndike & Lohman, 1990, p. 67). Indeed, he argued that it was not possible to construct an ability test that was unaffected by learning. Although he understood the intuitive appeal of separating "original capacity" from "whatever education [and experience] has added there to" (p. 95), twenty years of research convinced him that the idea was fundamentally flawed.

The impossibility of also purging ability from measures of achievement has received less attention, but is equally important. Again, E. L. Thorndike (1917; see also R. L. Thorndike, 1973-74) provides an example in a classic paper on "Reading as reasoning."

> Reading is a very elaborate procedure, involving a weighing of each of many elements of a sentence, their organization in the proper relations to one another, the selection of certain connotations and the rejection of others, and the cooperation of many forces to determine final responses. In fact…the act of answering simple questions about a simple paragraph…includes all the features characteristic of typical reasoning (E. L. Thorndike, 1917, p. 323).

For many, the transition to a new understanding of ability comes after realizing that what is measured by intelligence tests is not a fixed capacity but something very much the product of education and experience. J. M. Hunt recounts such an experience. His book *Intelligence and Experience* (1961) summarizes research on the effects of experience on the development of intelligence. In trying to explain why challenges to his belief in a fixed intelligence was so unnerving, Hunt appealed to Festinger's (1957) theory of cognitive dissonance:

> In his own professional life history, the writer finds in himself some evidence of [cognitive dissonance]. So long as he was professionally identified with the testing function, it was highly comforting to believe that the characteristics tested were fixed in individuals. Evidence hinting that these characteristics were not fixed produced intense dissonance, for it threatened his belief in fixity and the adequacy of his professional function as well. (pp. 14-15)

There are many other examples of this transition in beliefs about the relationship between conceptions of intelligence (or ability) and achievement. For most theorists, the changes are less well documented than in Hunt's case. But the products of these changes are clear, even though the transitions themselves are usually less transparent. Further, the diversity of theories individuals espouse increases as their expertise grows. In other words, development is best envisioned not as a ladder with a fixed sequence of steps, but rather as an inverted cone in which the range of possibilities increases as theories increase in sophistication. The progression goes something like this.

Level 1. Naïve nominalism or "Things are what they say they are." The person at this level believes that ability tests measure (or ought to measure) innate potential. This means that scores on an ability test should not be influenced by culture, education, personal experience, or motivation. Similarly, achievement tests measure (or ought to measure) only knowledge and skills learned in school. Performing better on an ability test than on an achievement test is interpreted as "not living up to one's potential." Virtually everyone starts with this understanding. Most retain this belief unless they encounter evidence that challenges it. Then either they acquire new beliefs or engage in various repair strategies to preserve their Level 1 beliefs.

Level 2. Understands Jangle fallacy: Ability and achievement tests seen as exchangeable. The person at this level has encountered and accepted evidence on the overlap between measures of intelligence and achievement. This evidence may come from statistics that show high correlations between ability and achievement test. Less formally, it may come from an inspection of ability and achievement tests that shows similarities in the content and structure of items on the two types of tests.

Reactions to this knowledge take several forms. Some look at the overlap and conclude that it reflects general ability (Jensen, 1998; Spearman, 1923). Others look at the same data and say that the overlap is mostly the product of learning (Ferguson, 1956; Humphreys, 1981; Thorndike, Bregman, Cobb, & Woodyard, 1926). Although both of these explanations account for the overlap between ability and achievement tests, neither explanation has much to say about the differences between them. A popular solution that attends both to similarity and difference is to envision a continuum in which tasks vary by their novelty. The more achievement-like tasks are placed at the low-novelty end of the continuum, whereas the more ability-like tasks are placed at the high-novelty end of the continuum. A continuum like this may be found in the writings of Stern (1914), Thorndike et al. (1926), Anastasi (1937), Cattell (1943, 1963), Cronbach (1970), Snow (1980), and Sternberg (1985), to name a few. Placing both types of tasks on the same continuum recognizes their commonality. Placing them at opposite ends of the continuum also recognizes their uniqueness.

Interpretations of this continuum vary. For example, in his original proposal of the theory of fluid and crystallized abilities, Cattell (1943) emphasized the equality of the two intelligences that he had distinguished. 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 constellation of crystallized abilities. Although Horn (see, e.g., Horn & Noll, 1997), Snow (1980), and others have repeatedly disputed this interpretation, it is the one that is most often presented in elementary textbooks that discuss the theory of fluid and crystallized abilities.

The attempt to interpret fluid reasoning abilities as the real intelligence is perhaps better understood as an attempt to preserve Level 1 beliefs in the face of Level 2 evidence. There are other common examples of this tendency. Most people who recognize that current tests of ability and achievement overlap assume that better tests could be constructed that would greatly reduce or eliminate the extent to which these tests call on similar knowledge and skills. An example was the hope expressed by many of us in the 1970s that we could construct better ability tests by directly measuring the higher-level cognitive processes people used when solving items on ability tests or other complex tasks (e.g., E. Hunt, Frost, & Lunneborg, 1973; Snow, 1978; Sternberg, 1977). A more extreme version of this view has recently been championed by those who claim that only nonverbal tests should be used to measure abilities. On this view, tests that measure reasoning abilities ought not to be contaminated by content or skills that would influence performance on an achievement test. Some advocate the use of nonverbal tests such as the Progressive Matrices (Raven, 1998) or the Nonverbal battery of the Cognitive Abilities Test (CogAT; Lohman & Hagen, 2001). Even these tests, though, are sometimes considered contaminated because the directions are given orally and thus use words. Correspondingly, advocates of this view argue that achievement tests ought not to measure anything that could be labeled *ability* (Naglieri & Ford, 2005).

Level 3. The swamp. Complications everywhere I look! Those who get beyond the idea that ability and achievement are separate things, or presently may be made to be separate things, confront a long list of further complications. Many who enter this swamp seem never to emerge from the other side. Rather, like marooned naturalists, they contentedly explore the wonders of a single island-domain, each of which is sufficiently complex to occupy a competent research team for an entire career. Examples include:

- *The effects of culture on cognition*. Beginning in the 1920s, some theorists noted that the very concept of intelligence is rooted in culture (see Anastasi, 1937; Degler, 1991). A culture-free measure of intelligence is thus something of an oxymoron. Similarly, what counts as achievement varies across cultures and eras. For example, handwriting, spelling, and computation skills are less valued today than 100 years ago. Independent thinking and problem solving are generally more highly valued.
- *The effects of education, practice, and training on abilities*. All abilities—from those required by the simplest reaction-time task to the most complex problem-solving task—respond to practice and training. Near the end of her career, Anastasi (1980) observed that much confusion could be avoided if the term *ability* were always prefaced by the adjective *developed*. Similarly, Snow observed that intelligence is not only education's most important raw material, but also its most important product (Snow & Yalow, 1982). Of course, this does not rule out a substantial role for biological factors in observed individual differences in ability at any point in the sequence of its development.
- *The effects of knowledge on thinking*. Just as people too glibly speak of the distinction between ability and achievement, many also speak too glibly about the separation of cognitive processes and the knowledge on which those processes act. One of the most important discoveries about human cognition is the extent to which thinking is bound to the objects of thought (Greeno, Collins, & Resnick, 1996). Put differently, there are no information-free cognitive processes. Reasoning does not exist as a module in the brain that can be applied like a tool to different problems (Lawson, 2004). Rather, how well we reason depends on how much we know. Language has particularly powerful effects on the development of thought, from the acquisition of simple perceptual concepts to complex assemblies of knowledge and skill that require many years to acquire.
- *The unity of the ability/achievement space*. If all abilities are achievements and all thinking is rooted in knowledge, then it makes little sense to talk about abilities and achievements as if they were qualitatively different things (Snow, 1980). Rather, many who study individual differences see a single space of developed competencies or abilities (Carroll, 1993; Cronbach, 1990; Horn & Noll, 1997; Humphreys, 1981). Some of these abilities are developed primarily through formal schooling, others through out-of-school experiences common to most children in a culture, and yet others through experiences that are unique to the individual. Different tests sample from different regions of that space of competencies. Achievement tests sample a broad range of knowledge and skills acquired primarily in school-like activities; ability tests such as the Otis-Lennon School Ability Test (Otis & Lennon, 1997) or the Cognitive Abilities Test (Lohman & Hagen, 2001) emphasize reasoning abilities that are required by and developed through experiences in- and outside of school. Most

individually administered ability tests such as the Stanford-Binet V (Roid, 2003) and especially the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001) sample a much broader array of abilities such as Comprehension-Knowledge, Long-term retrieval, Visual-Spatial ability, and Short-Term memory. Put another way, although general reasoning abilities depend on knowledge, what is known includes much more than those associations that facilitate reasoning. Reasoning abilities are thus a subset of the much larger domain of developed competencies.

- *The multidimensionality of the unified ability space*. For a very long time we have known that ability is a multidimensional, not unidimensional concept. Most theorists agree that that the 70+ abilities that have been identified can be organized in a hierarchy: a large *g* factor at the highest level, seven or more broad group abilities at the next level, and 50-87 primary abilities at the base (Carroll, 1993; McGrew, 2005). This theory is much less comforting to those who wish to make simple comparisons between ability and achievement than the quasi-hierarchical theory of Cattell (1963) that posited only two broad-group factors—fluid and crystallized ability. Fluid reasoning ability now includes verbal, quantitative, and figural aspects (Carroll, 1993); crystallized abilities no longer include quantitative achievements (McGrew, 2005). There is no easy bifurcation of abilities into the two camps of ability and achievement constructs. Rather, some argue that the most defensible way to view the continuum from crystallized achievements through more fluid applications to creativity is within domains of knowledge or skill (see Snow, 1981; Sternberg, 1998).
- *Gc as the real intelligence*? There has long been a bias among researchers that fluid intelligence (Gf) represents the real, biologically determined intelligence, whereas crystallized intelligence (Gc) better represents the products of investing this biological intelligence in particular experiences. Although there is some evidence that this may indeed be the case with very young children, thereafter “Gc may precede and do more to determine Gf than the reverse” (Horn & Blankson, 2005, p. 64; see also Lohman, 1993). Gf may also not be as central to models of abilities as many have believed. In his last published paper, Carroll (2003) reported a reanalysis of the Woodcock-Johnson (Revised) norm data. One of the major purposes of the analysis was to test Gustafsson’s (1988) hypothesis that Gf = g. The analysis, however, showed that the Gf factor was much less important than expected. The best measures of g were a vocabulary and mathematical problem solving test. As Carroll put it, “doubt is cast on the view that emphasizes the importance of a Gf factor.” Much other work shows that Gc abilities are better predictors of success both in the work place and in school than Gf abilities. Comparisons of the two become increasingly difficult, however, as children mature and acquire specialized knowledge. Measures of Gc sample an ever decreasing fraction of the domains in which individuals have developed expertise. The geography portion of the standard social studies test will not show the enormous geographical knowledge of the 13 year old who knows most of the major rivers in the world. Because they must be fair to all, standardized achievement tests measure what Horn and Blankson (2005) call dilettante knowledge of the culture. Attempts to develop more in-depth measures of domain-specific knowledge show enormous differences between the performance of high-school students and adults (Ackerman, 2000). But it is this expert knowledge and skill that is the real intelligence in the adult world (Hunt, 2000).
- *The impact of affect and volition on cognition*. Aristotle distinguished between cognition, affection, and conation—or knowing, feeling, and willing. Modern research on cognition shows that thinking is deeply enmeshed with affect. Interest (or disinterest), surprise (or boredom), enjoyment (or disgust) moderate what we remember about a topic, how deeply we think about it, and how long we will persist in thinking about it. Similarly, the choices we make as we embark on a task, or when we first encounter difficulty or distraction, also impact the success of our efforts. Put differently, one cannot estimate how well people think unless they are willing to try their best. Even then, they will generally do better if the topic interests them and if they feel that they are having success at it. Further, the knowledge and skill that they assemble both reflects and feeds into interest. There is no way to separate the measurement of ability from motivation or feeling.

- *The effects of experience on brain structures.* As with other dichotomies, the simple distinction between biology (or genetics) and experience does not survive close inspection. For example, even if one could somehow fulfill E. L. Thorndike's dream and measure the neuronal connections of the neonate's brain, the measure would not describe her biological structure for long. We now know that the brain is changed by experience. Extensive experience in a domain effects substantial changes in the structure of the brain and the way it processes information (Nelson, 1999). At a molar level, this means that the biological contribution to individual differences in ability is moderated by the quality of the environment in which the child is raised. Heritability is substantial for children from high SES families but much lower for children from low SES families (Turkheimer, Haley, Waldron, D'Onfrio, & Gottesman, 2003). Further, as Cronbach (1976) noted, virtually any statement made about the heritability of tests of general ability would apply with equal force to measures of academic achievement. This does not mean that anyone can do anything. A naïve environmentalism is surely more misleading than a naïve nativism. What it means is that, just as the neonate must grow physically to keep up with his peers, so too must he grow cognitively. The experiences that feed this growth determine subsequent cognitive status even more surely than nutritional intake determines subsequent physical stature.
- *The contextual specificity of thinking.* People who study individual differences generally examine only that portion of the variability in behavior that generalizes across tasks. However, what generalizes is typically only a small portion of the variability in items. Further, even when one combines scores across items, factors that are specific to the test (such as the format and sample of items) are as important as the ability dimension on which the test has its highest loading. For example, in one large study that included many ability and achievement tests, we found that scores on an hour-long version of the Raven Progressive Matrices loaded 0.7 on the general factor (Marshalek, Lohman & Snow, 1983). Squaring the factor loading gives the percent of variance on the test that it explains. Here this means that only about half of the variability in Raven scores was explained by g. Some of the non-g variance in the test could be attributed to other factors (particularly Gv). But most was specific to the test. This matters for two reasons. First, test users get both parts—the portion that generalizes and the portion that is unique. This means that even tests that load on the same factors will often give quite different scores to the same individual. Second, it means that context is more important in cognition than most ability theorists seem to appreciate.

Level 4. Systems theories. Given the scope and complexity of research on cognition, it is not surprising that very few scholars are able to envision theories or paradigms that integrate these diverse themes into a coherent whole. Two of the most impressive efforts are those of Robert Sternberg and Richard Snow.

Sternberg has been an incredibly prolific contributor to the field. Most students are familiar with his triarchic theory of intelligence (Sternberg, 1985, 1998). The theory has three subtheories: contextual, experiential, and componential. The contextual theory addresses the issue of which abilities a particular culture values as indicants of intelligence. Sternberg argues that the major indicants of successful intelligence in contemporary U.S. culture are analytic abilities, creative abilities, and practical abilities. The experiential theory addresses the question of the relative novelty of the task chosen as an indicant of intelligence for the person in question. It thus addresses the effect of experience on the development of abilities. Finally, the componential theory addresses the issue of the cognitive processes the person uses to solve the task. This links the theory to modern theories of human cognition.

Snow's theory is less well known, but in many ways is easier to apply to the problems educators face (for an introduction, see Corno, Cronbach, Kupermintz, Lohman, et al., 2002). The theory concerns how we might best design instruction to meet the needs of different learners (Cronbach & Snow, 1977). Turned around, it addresses the fundamental question of readiness to learn from particular set of educational activities (Snow, 1992, 1994; Snow & Lohman, 1984). The central construct is that of aptitude, by which Snow meant the degree of readiness to learn and perform well in a particular situation or domain. Aptitudes for learning are therefore tied both to what must be learned (i.e., What kind of expertise do we aim to develop?) and to the learning context (i.e., How are students expected to learn?) Students who will have a difficult time acquiring one type of expertise (e.g., mastering algebra) may have less difficulty acquiring expertise in another domain (e.g., creative writing). Those who might have difficulty succeeding under one instructional arrangement (e.g., large lecture class) might succeed more readily under another (e.g., computer-assisted instruction). Deciding which students are most likely to develop a particular type of expertise thus begins with a careful analysis of what constitutes expertise in the domain and how it is developed. Next, one looks at the demands and opportunities of the different educational paths offered for those who wish to develop expertise in the domain. What must students know and be able to do in order to succeed in each alternative route to the attainment of expertise? Aptitudes for learning typically include prior knowledge and skill in the domain, the ability to reason in the symbol system(s) used to communicate new knowledge, interest in the domain, and persistence. The theory thus turns the question of intelligence on its head. One begins not with a catalog of the person's standing on a given set of ability dimensions, but rather with a clear statement of where one wants to go (What kind of expertise?) and of the paths that will be made available to different learners to attain that expertise (What kinds of instruction?).

Physician, Heal Thyself

As I look at my own understanding of human abilities, I can see a slow and at times difficult progression through this sequence of beliefs. When I first learned about ability tests, I presumed that, when administered properly, they estimated a largely fixed characteristic we called *IQ*. Then, when I was in graduate school, I learned about the enormous overlap between tests of ability and tests of achievement. Because I was engrossed in research on human cognition, I assumed that the overlap reflected the fact that more able students learned more from their experiences than less able students. However, I was not convinced that existing measures of cognitive abilities were well grounded in psychological theory. This belief was shared by many others who were caught up in the cognitive revolution in psychology. We thought that we could develop better, purer measures of ability by isolating and then measuring particular cognitive processes. I was working under Richard Snow's direction. But Robert Sternberg was doing similar work under the direction of other faculty at Stanford. Earl Hunt and his students (at the University of Washington) and Robert Glaser and his students (at Pittsburgh) were also chasing the process phantom. We all eventually discovered (those studying artificial intelligence were surely there first) that process was inseparable from content (see Dehn & Shank, 1982). The information-free cognitive processes that E. Hunt et al. (1973) hoped to find never appeared. Indeed, by 2000, E. Hunt was arguing that crystallized intelligence (i.e. general achievement) was closer to the real intelligence than its much-ballyhooed partner—fluid intelligence. Sternberg (1985), Snow (1994), Glaser (1984), and many others had arrived at similar conclusions.

Now, after 30 years of research and teaching, I have come to appreciate not only the enormous complexity the field, but also the value of synthetic theories like those of Sternberg and Snow that go well beyond the level of understanding reached by even most professionals. Unfortunately, it is not at this level at which debates about the separability of ability and achievement are conducted. Indeed, most participants seem stuck at Level 2. Their arguments are in large measure designed to appeal to the even larger segment of the population whose experiences have not challenged them to move beyond Level 1.

The Enmeshment of Process in Knowledge

The assertion that ability should not be contaminated by achievement generally goes hand in hand with the belief that good ability tests should be culture free and should measure innate ability (for critiques of this view, see Anastasi & Urbina, 1997, or Scarr, 1994). The only way this is possible is if ability is not developed through experience and if tests could see through the veneer of culture, experience, and motivation. It also generally assumes that ability is something like a mental tool that can be applied at will to different intellectual content. Psychologically, it means that process can be separated from knowledge. However, not one of these beliefs has survived empirical scrutiny.

What is the alternative view? It is that there is a single universe of developed human competencies or, as Anastasi (1980) calls them, "developed abilities" (p. 3). Some tests sample broadly from that universe; some sample more narrowly. Some sample abilities that are developed through formal schooling; some sample abilities developed through experiences in the broader culture. But all cognitive tests measure developed abilities. There are no exceptions. Content and context are critical both for the development of ability and the expression of it.

Figure 1 shows one way to envision this common space of developed competencies. The figure shows a universe of cognitive tasks that is partitioned into verbal, quantitative, and spatial regions. Subsets of tasks that elicit or require reasoning abilities are highlighted. Although I focus on reasoning abilities, similar subsets of tasks could be noted for any of the other seven to 10 broad group or 56 primary abilities in Carroll's (1993, 2003) theory. There are three key ideas here. First, instead of two independent or overlapping circles (one for ability and the other for achievement) we now have one circle. Second, because the processes we call *reasoning* are invoked by many different types of tasks, they appear at multiple locations within the space. And third, tasks that require reasoning tend to be near the center of this circle, which is the location of the statistical variable called *g* in a two dimensional space (Marshalek, Lohman, & Snow, 1983).

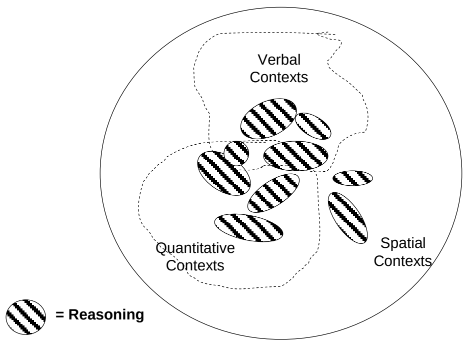

Figure 1. *The unified space of developed abilities. Abilities fall in three slices of the model: verbal, quantitative (symbolic), and spatial (figural) contexts. Tasks that emphasize reasoning are highlighted. G falls at the center of this perfectly balanced universe (see Marshalek, Lohman, & Snow, 1983). The subsets of tasks set off by dashed lines suggest samplings for verbal and quantitative achievement tests. These include tasks that require reasoning, and more.*

Notice that the ability to reason with figural/spatial stimuli, while also close to *g*, is also somewhat distal from the ability to reason with words or with quantitative concepts. Elsewhere, I have discussed the implications of this for identifying academically able students using nonverbal, figural reasoning tests (Lohman, 2005b). The issue is not a straightforward as most seem to believe. Probably the best way to understand both the advantages and limitations of such measures is by analogy to physical abilities. Like scores on a figural reasoning test, running speed is an important aptitude for many sports. It is also a measure of physical fitness that does not depend on access to specialized equipment or training. But the best swimmers, basketball players, or gymnasts will only sometimes excel at running. Thus, even though it is surely better than no selection test, a series of running competitions would not be the best way to identify those who currently display or who are most likely to develop excellence in particular sports.

Language and Thought

The use of language in testing has once again generated controversy, largely because of the linguistic diversity of children in the U.S. In the desire to be fair, and because of the erroneous belief that test scores for all children can only be interpreted by comparing them to the same age- or grade cohort, there are once again calls to evaluate the cognitive development of all children using tests that reduce as much as possible the contributions of language. Unfortunately, expelling language also expels an enormous amount of cognition. Words not only express thought but also give birth to new ways of thinking. Other symbol systems such as those used to communicate mathematical and musical concepts do the same.

Gentner's investigations of children's acquisition of concepts show why words are so important for the development of cognition (see Gentner & Goldin-Meadow, 2003). Words facilitate the ability of children to notice similarities and then derive concepts that capture regularities across situations. When adults repeatedly use the same word to describe different situations, the child is much more likely to notice the commonalties across the situations and *automatically* make an inference about the meaning of the concept. If the adult uses the word *yellow* to describe balls, cars, walls, and flowers, children are encouraged to notice this commonality and to learn a label that will enable them to elicit the concept both in their own thinking and in the minds of others. Without words and an adult culture guide, the child is aswim in a world in which hundreds of things change across time and place.

From the humble beginnings of a handful of concepts and words that can evoke them, children assemble vast networks of increasingly sophisticated and abstract concepts. A child's reasoning ability is not some innate function that is divorced from this knowledge base. As Bruner (1957) noted, reasoning is best seen in the ability to go beyond the information given. When viewed in this way, *reasoning* describes a particular collection of methods for accessing and using one's knowledge to leverage new understanding. Concretely, then, *one cannot measure the sophistication of a children's reasoning or problem solving abilities unless one presents problems that allow children the opportunity to make good use of the knowledge that they have assembled.* The horns of the dilemma are these: On the one hand, unfamiliar language can be a stumbling block; on the other hand, to pare the world down to a small set of regular geometric shapes eliminates most of what children know and can do.

Braden (2000) has discussed this issue in several places, perhaps most succinctly in a special issue of the *Journal of Psychoeducational Assessment* that was devoted to the topic of nonverbal testing. As he sees it, advocates of nonverbal testing recognize the fact that tests that use language and rely on verbal knowledge may introduce construct-irrelevant variance into the testing session for those students who do not speak the language of the test, who are deaf, or who have other specific linguistic disabilities. However, he also notes that those who advocate nonverbal tests are often insensitive to construct underrepresentation—which is the other primary threat to test score validity. In other words, for children who speak a language (especially the language used in the test) or who have attended school and learned some of basic mathematical concepts, not using words or quantitative concepts prohibits them from showing how well they can reason. This is as true for minority students as for nonminority students. For the vast majority of these children, the problem is not that the test uses words or quantitative concepts, but which norm groups are used to interpret the scores they obtain that best support the inferences that are made (see Lohman, in press-a, in press-b).

Measuring Reasoning Abilities

If reasoning abilities are best measured within contexts that allow students to make good use of their knowledge, then can reasoning and knowledge be distinguished at all? The first step is to understand that reasoning is not a cognitive module that is applied to a knowledge base. Rather, it describes a particular way of accessing and using that knowledge to further ones' understanding or problem solving. Reason abilities in a domain are thus a subset of a larger set of abilities that collectively index what one knows and can do in that domain. E. L. Thorndike's example of the relationship between reading and reasoning provides a good example of how verbal reasoning abilities can form part of a broad achievement construct s such as reading ability. Good reading tests sample broadly from the many different things that a skilled readers do while reading. These include decoding familiar and unfamiliar words, coordinating the ideas extracted from the text with those extracted from a picture or diagram, building and coordinating internal mental models that represent the ideas in a text and a visual image of them, activating relevant knowledge in long term memory and establishing new associations with ideas gleaned from the text, inferring the main idea of the text, responding with amusement or annoyance at the content of the article or the style of the writing, and much more. Reasoning processes are centrally involved in several—but not all—of these activities. Reading requires much more than verbal reasoning, but verbal reasoning is essential for that part of reading we call comprehension. A test of verbal reasoning abilities, then, would try to make salient the search, retrieval, and comparison processes that together constitute reasoning while reducing the influence of other reading processes. Like the flower and the plant, the two cannot be separated. But it is possible to clear the undergrowth and prune the plant so that the flower stands out clearly.

This is why artificial item types such as analogies or classification problems are so useful. They make salient the reasoning process and, when items are well constructed, reduce the impact of other processes. Further, thirty years of research on the psychology of such test-tasks can greatly facilitate efforts to build reasoning tests with them that enhance the contribution of construct-relevant sources of task difficulty while minimizing potential contributions of construct-irrelevant sources of difficulty (Lohman, 2005a; Sternberg, 1985).

How can these tasks be constructed to emphasize reasoning? Vocabulary tests such as those used on the Wechsler (2003) and Stanford-Binet (Roid, 2003) scales provide excellent examples. Children infer the meanings of words by remembering and attending to similarities and differences in the way words are used in different contexts. For measuring reasoning, the critical factor is not students' knowledge of infrequent words but rather the precision of their understandings of relatively common, but abstract words (Snow & Lohman, 1989). Knowledge of infrequent or specialized words is actually a better measure of interest and specific achievement. Having only a vague understanding of the connotations of commonly used words is one of the oldest and most commonly observed characteristics of less able students (R. M. Thorndike & Lohman, 1990).

The sentence completion test provides another example. This is probably the oldest and still one of the best formats for measuring verbal reasoning abilities (Ackerman, Beier, & Bowen, 2000). It focuses on the central task verbal comprehension: making inferences about the meaning of unknown words or incompletely spoken language. Good inferences both honor the syntactic constraints in the sentence and make full use of the conceptual information in it. As with the vocabulary test, however, the abilities measured by the task depend on the nature of the inferences the examinee must perform in order to fill in the blanks. For example, most third graders can easily read the sentence *Cats have two eyes but only one* ____. However, some select a foil such as *ears* because they make inferences associatively rather than logically (see also Lohman, 2005a, p. 118).

Empirical Evidence

The statistical procedures one might employ to test the assertion that ability and achievement tests measure the same thing have been around for as long as the issue itself. Kelley (1927) understood that reliability of the separate measures was critical. However, he used these reliabilities to estimate the percent of systematic or true variance that the two measures shared. R. L. Thorndike (1963) emphasized instead the reliability of the non-shared variance. In a classic book, he addressed the criteria one might use to decide whether ability and achievement tests measure sufficiently different things to warrant measuring both.

Suppose we predict a particular aspect of achievement such as the Reading Total score on the Iowa Test of Basic Skills (ITBS; Hoover, Dunbar, & Frisbie, 2001) from scores on the CogAT verbal reasoning battery. This produces two scores for each child: the observed ITBS Reading Total score and a predicted reading score. There are two critical issues here. First, what is the magnitude of the correlation between these two scores? If the correlation is low (say r = .5 or less), then the reasoning test is not of much use for making predictions about the reading achievement for individual students. Psychologically, this says that it is unlikely that the sorts of abilities measured by the reasoning test are important aptitudes for the development of the reading skills measured by the achievement test. If, on the other hand, the correlation is extremely high (say r = .9), then differences between observed reading achievement and predicted reading achievement may simply reflect errors of measurement in the ability and achievement tests. The horns of the dilemma, then, are defined by the need to measure reasoning abilities in the contexts that are important for reading, but not to duplicate the reading test.

How well does the reasoning score estimated from these tasks predict reading abilities? The median correlation (grades 3-8) between CogAT Verbal and ITBS Reading Total is r = .82 (Lohman & Hagen, 2002). This says that the CogAT Verbal score has excellent predictive validity for ITBS Reading scores. But how dependable are the differences between observed and predicted achievement? Actually, difference scores are notoriously unreliable. This is because when one subtracts two scores, any systematic variance shared by the two tests (which is given by the square of the correlation between them) is eliminated. In the example above, this is $.82^2$ or 67% of the variance in each test. To make matters worse, errors of measurement do not cancel, but are both represented in the difference score. *Thus, a critical issue for the comparison of two scores is the reliability of each*. This is one of the main reasons why CogAT measures each of three reasoning abilities in multiple formats with multiple items. The result is a set of reasoning test scores that are highly reliable (e.g., average $r_{xx'}$ = .95 for verbal scores on the multilevel battery). This means that the reliability of the difference between observed and predicted achievement are also quite high. In the case in which ITBS Reading Total scores are predicted from the CogAT Verbal battery, the reliability of difference score is $r_{xx'}$ = .76 (Hoover, Dunbar, & Frisbie, 2003). If the two tests measured the same abilities, the reliability of this difference score would be zero.

Implications for Identification and Instruction

The process of identifying children who are likely to excel academically needs to be grounded in something more serious than a personal preference for one kind of measure over another. Ultimately, the goal is to identify and then provide developmentally appropriate instruction for (a) students who currently display extraordinary levels of accomplishment in domains for which instructional assistance can be provided and (b) students who show promise for developing this sort of academic excellence. Importantly, the types of academic support that can be offered constitute an important part of the selection problem. Children who excel in writing narratives often do not excel in mathematical reasoning. Children who can succeed when working in concert with others may not succeed when working alone. In other words, defining the treatment is part of defining the aptitude (Snow & Lohman, 1984).

Those who take the time to engage seriously in thinking about the specific demands of different learning tasks and instructional environments will discover why understanding readiness (or aptitude) is generally much more helpful than trying to measure intelligence, however defined. Research on academic learning shows that the best predictors of subsequent learning in a domain are current achievement in that domain, the ability to reason in the symbol systems of that domain, interest in the domain, and the ability to persist in striving to attain excellence (Corno et al., 2002). The relative importance of each factor varies somewhat with age and experience. For example, general reasoning abilities in the symbol systems of the domain are more important for novices whereas prior achievement becomes more important as students acquire expertise. Thus, selection policies need to be sensitive to these developmental differences.

Elsewhere (Lohman, 2005b, in press-a, in press-b), I have used an aptitude perspective to show how schools can better identify those who either currently display academic excellence and those who do not display excellence but are most likely to develop it if given additional instructional resources. This distinction is especially helpful in identifying academically talented minority students. For example, one can usefully ask whether an ELL student can read or can reason with English words without making inferences about his or her verbal ability in any grander sense. Rather the goal would be to determine if the student displays sufficient readiness to learn in a classroom in which English is the language of instruction. Larger inferences about potential to learn require that the student's performance be compared with others who have had similar opportunities to acquire the knowledge and skills measured by the test, not just with all other students in the norm group who share the same birth date or school grade. This is not difficult to do. If competence is low, then one can intervene to improve those skills in reading and reasoning with English that are critical aptitudes for classroom learning. Note that a much lower level of competence in the English language would be required to succeed in most mathematics classes, although even here low levels of oral and reading skills can severely limit classroom participation.

Understanding that all abilities are both grounded in biological processes yet very much developed has broader implications for programs for the academically gifted. The best programs for academically gifted children see their mission as developing talent—not merely discovering it. Programs might better communicate this goal to the public if they emphasized more their role in developing academic excellence and spoke less about giftedness. Anyone can aspire to excellence. Giftedness, however, has connotations of fixedness that are rightly resented by those who score lower on tests that measure the abilities and achievements used to define the construct. Saying that a child comprehends poorly what she reads has different instructional consequences than saying that she has a low WISC Verbal IQ. Yet, as I have noted, the two scores overlap much more than they differ. The problem, then, is not the measure but the label. Both public and professionals so seriously misunderstand IQ scores that it would be a great boon if the field could begin to define excellence in ways that do not depend on scores that are reported in this metric.

Conclusions

In one of his more memorable quotes, H. L. Mencken once observed that for every complex human problem there is always an easy solution that is neat, plausible, and wrong. There are two easy, neat solutions to the problem of overlap between ability and achievement tests. First, measure only the overlap, and, depending on one's predilections, ascribe it either to ability or to achievement. Second, attempt to eliminate words, numbers, or anything else explicitly taught in school from the measurement of ability (and, correspondingly, anything that looks like ability from the measurement of achievement). Both of these solutions are wrong. Each hopes to transform an inherently ill-structured problem for which there is no perfect answer into a well-structured problem with an all-or-nothing solution. Unfortunately, understanding why a more nuanced judgment is necessary requires more expertise about the nature and nurture of human intellectual competence than most of us can achieve without many years of study. The path that leads from a theory of naïve nominalism to the inclusive systems theories of Sternberg (1998) or Snow (1994) has many stopping off points. Not everyone has the time or inclination to attain a comprehensive view of the field. Nevertheless, a willingness to move beyond one's current beliefs—especially to discover what might be wrong with them—has always characterized more thoughtful professionals in all domains. As in other areas of science, one of the first steps in this process is to articulate one's beliefs. The second and harder step is to look closely at the disconcerting evidence that challenges rather than reinforces intuitions. The first goal of this paper was to provide a few such challenges for those willing to entertain them. The second goal was to describe something of the rewards of the journey ahead for those willing to embark on it.

References

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences, 55B* (2), 69-84.

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2000). Explorations of crystallized intelligence: Completion tests, close tests, and knowledge. *Learning and Individual Differences, 12,* 105-121.

Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

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

Anastasi, A. (1980). Abilities and the measurement of achievement (pp. 1-10). In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade: Proceedings of the 1979 ETS Invitational Conference.* San Francisco: Jossey Bass. (*New directions for testing and measurement* series)

Assouline, S. (2003). Psychological and educational assessment of gifted children. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 124-145). Boston: Allyn & Bacon.

Braden, J. P. (2000). Perspectives on the nonverbal assessment of intelligence. *Journal of Psycho-educational Assessment, 18,* 204-210.

Bruner, J. (1957). Going beyond the information given. In J. S. Bruner (Ed.), *Contemporary approaches to cognition* (pp. 41-69). Cambridge, MA: Harvard University Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: A tribute to Arthur Jensen* (pp. 5-21). Oxford, UK: Elsevier.

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.

Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., et al. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow.* Hillsdale, NJ: Erlbaum.

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

Cronbach, L. J. (1976). Measured mental abilities: Lingering questions and loose ends. In B. D. Davis & P. Flaherty (Eds.), *Human diversity: Its causes and social significance* (pp. 207-222). Cambridge, MA: Ballinger.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

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

Degler, C. (1991). *In search of human nature: The decline and revival of Darwinism in American social thought.* New York: Oxford University Press.

Dehn, N., & Shank, R. (1982). Artificial and human intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 352-391). New York: Cambridge University Press.

Duit, R. (1991). Students' conceptual frameworks: Consequences for learning science. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 65-85). Hillsdale, NJ: Erlbaum.

Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology, 10,* 121-131.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Roco, Peterson.

Gagné, F. (2003). Transforming gifts into talents: The DMGT as a developmental theory. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 60-74). Boston: Allyn & Bacon.

Gardner, H. (2003, April). Multiple intelligences after twenty years. Paper presented at the meeting of the American Educational Research Association, Chicago. Retrieved August 25, 2005, from Harvard University, Project Zero Web site: http://www.pz.harvard.edu/PIs/HG.htm

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought.* Cambridge, MA: MIT Press.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Macmillan.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35-71). Hillsdale, NJ: Erlbaum.

Hofer, B. K., & Pintrich, P. R. (Eds.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing.* Mahwah, NJ: Erlbaum.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Test of Basic Skills: Form A.* Itasca, IL: Riverside.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2003). *ITBS Forms A & B guide to research and development.* Itasca, IL: Riverside.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding cognitive abilities In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 41-68). New York: Guilford Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

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.

Hunt, E. B. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12,* 123-130.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 87-122). New York: Academic Press.

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

Jensen, A. R. (1998). *The* g *factor: The science of mental ability*. Westport, CT: Praeger.

Kelley, T. L. (1927). *Interpretation of educational measurements*. Chicago: World Book.

Kelley, T. L. Ruch, G. M., & Terman L. M. (1922). *The Stanford Achievement Test*. Younkers-on-Hudson, NY: World Book.

Lawson, A. E. (2004). Reasoning and brain function. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 12-48). Cambridge, UK: Cambridge University Press.

Lohman, D. F. (1993). Teaching and testing to develop fluid abilities. *Educational Researcher, 22,* 12-23

Lohman, D. F. (2005a). Reasoning abilities. In R.J. Sternberg, J. Davidson, & J. Pretz (Eds.), *Cognition and intelligence* (pp. 225-250). New York: Cambridge University Press.

Lohman, D. F. (2005b). The role of nonverbal ability tests in the identification of academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49,* 111-138.

Lohman, D. F. (in press-a). An aptitude perspective on talent: Implications for the identification of academically gifted minority students. *Journal for the Education of the Gifted.* Retrieved August 25, 2005, from The University of Iowa, College of Education Web site: http://faculty.education.uiowa.edu/dlohman/

Lohman, D. F. (in press-b). *Identifying academically gifted minority students: A user's guide to theory and method.* (Research Monograph). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved August 25, 2005, from The University of Iowa, College of Education Web site: http://faculty.education.uiowa.edu/dlohman/

Lohman, D. F., & Hagen, E. (2002). *Cognitive Abilities Test (Form 6): Research handbook.* Itasca, IL: Riverside.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' Objectively Determined and Measured." *Journal of Personality and Social Psychology, 86,* 96-111.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107-127.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 136-181). New York: Guilford Press.

Naglieri, J. A., & Ford, D. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly, 49,* 29-36.

Nelson, C. A. (1999). Neural plasticity and human development. *Current Directions in Psychological Science, 8,* 42-45.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Mahwah, NJ: Erlbaum.

Otis, A. S., & Lennon, R. T. (1997). *Otis-Lennon School Ability Test* (7th ed.). San Antonio, TX: Harcourt Brace.

Raven, J. C. (1998). *Manual for Raven's progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Scarr, S. (1994). Culture-fair and culture-free tests. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 322-328). New York: Macmillan.

Snow, R. E. (1978). Theory and method for research on aptitude processes. *Intelligence, 2,* 225-278.

Snow, R. E. (1980). Aptitude and achievement (pp. 39-60). In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade: Proceedings of the 1979 ETS Invitational Conference.* San Francisco: Jossey Bass. (*New directions for testing and measurement* series).

Snow, R. E. (1981). Toward a theory of aptitude for learning: Fluid and crystallized abilities and their correlates. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 345-362). New York: Plenum.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist, 27,* 5-32.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3-37). Cambridge, UK: Cambridge University Press.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347-376.

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.

Snow, R. E., & Yalow, E. (1982). Education and intelligence. In R .J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493-585). Cambridge, UK: Cambridge University Press.

Spearman, C. E. (1923). The nature of intelligence and the principles of cognition. London: Macmillan.

Stern, W. (1914). The psychological methods of testing intelligence. *Educational Psychology Monographs, 13*. (G. M. Whipple, trans.)

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

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1998). Applying the triarchic theory of human intelligence in the classroom. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: Theory into practice.* Mahwah, NJ: Erlbaum.

Tannenbaum, A. J. (2003). Nature and nurture of giftedness. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 45-59). Boston: Allyn & Bacon.

Thorndike, E. L. (1903). *Educational psychology*. New York: Lemcke & Buechner.

Thorndike, E. L. (1917). Reading as reasoning: a study of mistakes in paragraph reading. *Journal of Educational Psychology, 8,* 323-332.

Thorndike, E. L., Bregman, E. O., Cobb, M. V., & Woodyard, E. (1926). *The measurement of intelligence.* New York: Columbia University, Teachers College, Bureau of Publications.

Thorndike, R. L. (1963). *The concepts of over- and under-achievement*. New York: Columbia University, Teachers College.

Thorndike, R. L. (1973-74). "Reading as reasoning." *Reading Research Quarterly, 2,* 135-147.

Thorndike, R. M., & Lohman, D. F. (1990). *A century of ability testing*. Chicago: Riverside.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14,* 623-628.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson I II*. Itasca, IL: Riverside.