

An Aptitude Perspective on Talent: Implications for Identification of Academically Gifted Minority Students

David F. Lohman¹

The University of Iowa

Journal for the Education of the Gifted (in press)

Abstract

In this paper, I discuss the identification of academically gifted children from the perspective of aptitude theory. Aptitude refers to the degree of readiness to learn and to perform well in a particular situation or domain. The primary aptitudes for academic success are (1) prior achievement in a domain, (2) the ability to reason in the symbol systems used to communicate new knowledge in that domain, (3) interest in the domain, and (4) persistence in the type of learning environments offered for the attainment of expertise in the domain. Careful attention to the demands and affordances of different instructional environments enables educators to identify those individuals who are most ready to succeed in them. Although the principles discussed here are useful for all students, they are particularly important for the identification of academically promising minority students.

The goals of this paper are threefold. First, I offer a brief introduction to recent developments in the psychology of aptitude. Second, I show how the concept of aptitude can help clarify the goals that guide attempts to identify gifted students, the procedures that achieve these goals, and the sorts of research evidence that would support the process. Finally, I show how these concepts can assist in the identification of academically gifted students from underrepresented minority populations. In a nutshell, my argument is that (1) admission to programs for the gifted should be guided by evidence of aptitude for the particular types of advanced instruction that can be offered by schools; (2) the primary aptitudes for development of academic competence are current knowledge and skill in a domain, the ability to reason in the symbol systems used to communicate new knowledge in the domain, interest in the domain, and persistence; (3) inferences about aptitude are most defensible when made by comparing a student's behavior to the behavior of other students who have had similar opportunities to acquire the skills measured by the aptitude tests; however,

(4) educational programming and placement should be based primarily on evidence of current accomplishment.

Taken together, these claims have several policy implications. The first implication is that there are—conceptually at least—two groups of children who should be considered when designing programs for the academically gifted. The first group consists of those students who currently display academic excellence in a particular domain. To facilitate discussion, I will refer to these students as belonging to the *high-accomplishment* group. Although the measurement of academic accomplishment is not a trivial matter, these students are generally easier to identify than those in the second group. Students in the second group do not currently display academic excellence in the target academic domain but are likely to do so if they are willing to put forth the effort required to achieve excellence and are given the proper educational assistance. I refer to these students as belonging to the *high-potential* group. Students commonly fall in the high potential group because, through age, circumstance, or choice, they have not developed expertise in a particular domain. For example, if we define scholarly productivity or artistry in a domain as something beyond expertise (Subotnick & Jarvin, in press), then even the most accomplished children will at best exhibit high potential. If, on the other hand, *expertise* is defined in terms of reading or mathematical problem-solving skills well in advance of age or grade peers, then many more children will exhibit high accomplishment. However, some students who do not display high accomplishment might also currently do so, had they had the opportunities to develop these skills. Put differently, high-potential students display the aptitude to develop high levels of accomplishment offered by a particular class of instructional treatments.

The second policy point is that high-accomplishment students typically need different educational programs than high-potential students. Both groups need instruction that is geared to their current levels of accomplishment. Because their levels of accomplishment differ, instruction aimed at one group will often be inappropriate for the other group. An undifferentiated label such as “gifted” does not usefully guide educational programming for a group that contains a mix of both high-accomplishment and high-potential students.

The third point is that the distinction between high-potential and high-accomplishment students is critical in the identification of academically talented minority students. Many of the most talented

¹ This paper is based on an invited presentation at the Seventh Wallace National Research Symposium on Talent Development, Iowa City, IA, May 2004.

minority students will not have had opportunities to develop high levels of the skills valued in formal schooling. Therefore, identification of such students depends on a clear understanding of how one measures academic aptitude. The purpose of this paper is to offer some suggestions on how to do this.

Current Practices

How can we best identify academically gifted children? Should it be on the basis of an individually administered intelligence test, group-administered achievement test, or indices such as grades that are based on teacher judgments? Can we rely on test of creativity; a test of practical intelligence; or a nonverbal test, especially one that purports to be “culture fair”? What if we administered one or more performance assessments in different domains? If we use multiple procedures, will they be exchangeable, or should we array them in a matrix? If information is to be combined, how should we combine it in order to make good selection decisions? (For overviews, see Assouline, 2003; Hagen, 1980.)

One way to define intellectual giftedness is to catalog the ways in which individuals differ in cognitive abilities and achievements. The advantage of this approach is that there is now considerable consensus on number and organization of human cognitive abilities. The Cattell-Horn-Carroll (CHC) theory is probably the best current summary. It contains a three-level hierarchy: a general factor (G); eight to 10 broad group factors; and from 60 to 75 primary ability factors at the base (McGrew & Evans, 2004; Traub & McGrew, 2004).²

Oddly, many who acknowledge this model act as if it has only one factor—i.e., G—rather than 70 or 80. Surely G is important. Indeed it is the single most important factor in the model. But it is not the only factor. Furthermore, it is only the best predictor of academic success when measures of achievement are also aggregates over many different kinds of outcomes for many different courses of study. Put differently, G is a good predictor of undifferentiated outcomes. But once school achievements are differentiated in some way, then more differentiated prediction is needed. For example, if the criterion is competence in writing and speaking one’s native language, then tests of verbal reasoning and verbal fluency add importantly to the prediction of success.

² Following Snow & Lohman (1984) and Carroll (1993), I use the symbol *G* rather than *g* to denote the general factor in a representative battery of mental tests. This acknowledges the general factor without some of the interpretive entanglements that often accompany the factor Spearman dubbed *g*.

Tests of writing and speaking skills add even more. If the criterion is facility in acquiring a second language, other verbal abilities enter the mix. Similarly, if the competence is in mathematics or architecture or mechanical engineering, then yet other abilities add to the prediction of success afforded by G (Gustafsson & Baulke, 1993; Shea, Lubinski, & Benbow, 2001).

This immediately suggests that we are not interested in ability for ability’s sake, but in ability *for* something. We are not interested in identifying bright kids in order to congratulate them on their choice of parents or some other happenstance of nature or nurture. Rather, the goal is to identify those children who either currently display or who are likely to develop excellence in the sorts of things we teach in our schools. Identifying such students is a much more tractable problem than identifying all the ways in which people differ and then creating programs that will help individuals develop those many and varied gifts. Put differently, those who take an ability-centered approach to the identification of giftedness have no basis other than parsimony for designating one ability as more important than another ability. For example, it is only when we add the criterion of utility that general crystallized abilities become much more important than general spatial or general memory abilities in the identification of academic giftedness because crystallized abilities better predict school achievement, even though general crystallized, spatial, and memory abilities have equal stature in the CHC theory of human abilities. Additionally, the ability-centered approach offers no principled way for incorporating motivation, creativity, or any of the other factors we may think important into the selection process. Indeed, the MENSA Society is the example *par excellence* of the ability-centered approach to the identification of giftedness.

The first point, then, is that academic giftedness is best understood in terms of aptitude to acquire the knowledge and skills taught in schools that lead to forms of expertise that are valued by a society. We are interested in ability tests only because they help identify those who may someday become excellent engineers, scientists, writers, etc. In other words, we are interested in abilities because they are indicants of aptitude. They are not the only indicants, but one important class of indicants.

A Definition of Aptitude

So, what do I mean by *aptitude*? Although often rooted in biological predispositions, it is not something that is fixed at birth. Achievements commonly function as aptitudes—for example,

reading skills are important aptitudes for school learning. Indeed, aptitude encompasses much more than cognitive constructs such as ability or achievement. Persistence is an important aptitude in the attainment of expertise. Also, aptitudes are not necessarily positive. Some people have a propensity to have or to cause accidents; others to lie; others to be unsociable or even hostile. The intuitive appeal of theories of emotional intelligence is rooted in the common observation that a productive and happy life requires more than abrasive intelligence. Finally, and most importantly, the term *aptitude* does not refer to a personal characteristic that is independent of context or circumstance. Indeed, *defining the situation or context is part of defining the aptitude*. Changing the context changes in small or large measure the person characteristics that influence success in that context.

Aptitude is thus inextricably linked to context. Consider formal schooling. Students approach new educational tasks with a repertoire of knowledge, skills, attitudes, values, motivations, and other propensities developed and tuned through life experiences to date. Formal schooling may be conceptualized as an organized series of situations that sometimes demand, sometimes evoke, or sometimes merely afford the use of these characteristics. Of the many characteristics that influence a person's behavior, only a small set aid goal attainment in a particular situation. These are called aptitudes. Formally, then, aptitude refers to *the degree of readiness to learn and to perform well in a particular situation or domain* (Corno, Cronbach, Kupermintz, Lohman, Mandinach, Porteus, & Talbert, 2002). Thus, of the many characteristics that individuals bring to a situation, the few that assist them in performing well in that situation function as aptitudes. Those that impede their performance function as inaptitudes. Examples of characteristics that commonly function as academic aptitudes include the ability to comprehend instructions, to manage one's time, to use previously acquired knowledge appropriately, to make good inferences and generalizations, and to manage one's emotions. Examples of characteristics that function as inaptitudes include impulsivity, high levels of test anxiety, and prior learning that interferes with the acquisition of new concepts and skills.

Sometimes the same situation that elicits modes of responding that function as aptitudes can also elicit modes of responding that thwart goal attainment. For example, discovery-oriented or constructivist approaches to learning generally succeed better than more didactic approaches with more able learners (Cronbach & Snow, 1977; Snow & Yalow, 1982).

Ill-structured learning situations afford the use of these students' superior reasoning abilities, which thus function as aptitudes. However, anxious students often perform poorly in relatively unstructured situations (Peterson, 1977). Thus, the same situation that affords the use of reasoning abilities can also evoke anxiety. Recent efforts to understand how individuals behave in academic contexts have emphasized the importance of these clusters of traits that combine to produce the outcomes that we observe (Ackerman, 2003). Lubinski and Benbow (2000) have argued for the same sort of attention to diversity in the needs academically gifted students. Indeed, gifted students will vary as much from each other on those dimensions not correlated with G as students in the general population.

Measuring Aptitude

Aptitude is commonly inferred in two ways. In the first, we attempt to identify other tasks that require similar cognitive processes and measure the individual's facility on those tasks (Carroll, 1974). For example, phonemic awareness skills that facilitate early reading in Spanish for Hispanic students also facilitate early reading in English for these students (Lindsey, Manis, & Bailey, 2003). Thus, one can estimate the probability that Spanish-speaking students will learn to read English by measuring their phonemic awareness skills in Spanish. Similarly, dance instructors screen potential students by evaluating their body proportions, their ability to turn their feet outwards, and their ability to emulate physical movements (Subotnick & Jarvin, in press). Although none of these characteristics require the performance of a dance routine, all are considered important aptitudes for acquiring dance skills.

In the second way, aptitude is inferred from the speed with which the individual learns the task itself. Aptitude for a task is inferred retrospectively when a student learns something from a few exposures to that task that other students learn only after much practice. Indeed, the concept of aptitude was initially introduced to help explain the enormous variation in learning rates for different tasks exhibited by individuals who seemed similar in other respects (Bingham, 1937).

Understanding which characteristics of individuals are likely to function as aptitudes begins with a careful examination of the demands and affordances of target tasks and the contexts in which they must be performed. This is what we mean when we say that defining the situation is part of defining the aptitude (Snow & Lohman, 1984). The affordances of an environment are what it offers or makes likely or makes useful. Placing chairs in a

circle affords discussion; placing them in rows affords attending to someone at the front of the room. Discovery learning often affords the use of reasoning abilities; direct instruction often does not. Aptitude is thus linked to context. Unless we define the context clearly, we are left with distal measures that capture only some of the aptitudes needed for success.

An example may help. Selecting students for advanced instruction in science or literature using a measure of G is like selecting athletes for advanced training in gymnastics or basketball using a measure of general physical fitness. Many who display high levels of physical fitness would not have much skill or interest in either of these domains. Furthermore, particular aptitudes loom large in the development of high levels of competence. For example, those who succeed in gymnastics tend to have different physical characteristics than those who succeed in basketball (Tanner, 1965). More importantly, even though a distal measure, such as overall physical fitness, may work with tolerable accuracy in the entire population, it will fail abysmally in identifying the high achievers in particular domains.

The Non-exchangeability of Measures

There is much confusion about this in the educational literature, abetted in large measure by a misunderstanding of how to interpret correlations. Simply put, the fallacy is that if measures are highly correlated, one would identify more or less the same individuals on either measure.

Table 1 shows why this is not the case. The data come from the 2000 joint national standardization of Form A of the Iowa Tests of Basic Skills® (ITBS®, Hoover, Dunbar, & Frisbie, 2001) and Form 6 of the Cognitive Abilities Test™ (CogAT®, Lohman & Hagen, 2001a). Data are reported for grades 3 through 6 to give some idea of the extent to which patterns replicate across grades. Sample size is approximately 14,000 students per grade.

The question was whether highly correlated selection tests would all identify students who show excellent achievement in a particular domain. Consider reading abilities as an example. What percentage of the students who scored in the top 3% of the distribution of Reading Total scores (Reading Vocabulary plus Reading Comprehension) would we identify using a series of other selection measures? These measures are roughly ordered by their proximity to the ITBS Reading Total Score. They are

1. *ITBS Reading Total.* This is the criterion measure. By definition we would identify all of the students who score in the top 3% of the distribution of Reading Total scores.

2. *ITBS Composite.* Many schools use the Composite Score across all subtests of the ITBS to identify academically gifted children. But what percent of the best readers would be missed using this score? Reading comprehension is not only a critical aptitude for success on other subtests of the ITBS, but the Reading Total Score also enters into the computation of the ITBS Composite (so there is a statistical confounding as well). The median within-grade correlation between the Reading Total and Composite scores was $r = .91$ in this sample.

3. *CogAT regression estimate of ITBS Reading Total.* Here we based selection on a regression estimate of Reading Total from the three CogAT battery scores at each grade. The median weights were (.684) CogAT Verbal Battery + (.126) CogAT Quantitative Battery + (.056) CogAT Nonverbal Battery. The median within-grade correlation between this regression estimate and Reading Total scores was $r = .83$.

4. *CogAT Composite.* In addition to the three battery scores, CogAT reports a Composite Score. It is the best estimate of G on the CogAT. It is obtained by averaging the examinee's scale scores across the three batteries—i.e., (1.0) CogAT V + (1.0) CogAT Q + (1.0) CogAT N. The median correlation between the CogAT Composite and ITBS Reading Total scores was $r = .79$.

5. *CogAT Verbal Battery.* Verbal reasoning abilities are critical in the acquisition of both reading comprehension skills and reading vocabulary. Because of this, one might expect the CogAT Verbal Battery Score to predict reading abilities about as well as either the regression composite (variable 3) or the unit-weighted composite (variable 4). Its within-grade correlation with ITBS Reading Total was $r = .82$.

6. *CogAT Nonverbal Battery.* Some schools use nonverbal reasoning to identify gifted students. Although this is surely the most distal battery studied, its median correlation with Reading Total was still substantial (median $r = .62$).

Although there is some variation across grades, the row in Table 1 that reports the average percentage of the top readers identified by each measure nicely summarizes the data. Slightly more than half (54%) of the best readers would be identified if one used the ITBS Composite Score rather than the Reading Total Score. Put the other way, selection using the ITBS Composite Score would miss about half of the best

readers. This is not what most people would expect for two variables that correlate $r = .91$.

Using the best linear combination of CogAT scores gets 36% of the best readers, which is about the same as the percentage that would be identified using the CogAT Verbal Battery Score alone (35%). The CogAT Composite Score gets only 32%. And the Nonverbal Battery identifies only 18% of the best readers. Table 2 shows a parallel set of analyses on the ITBS Mathematics Total Score.

Clearly, different measures do not identify the same students in spite of the fact that they are highly correlated. In part, this is because correlations generally imply far less agreement between scores than most people think, especially for extreme scores

(see Lohman, 2004, for examples). There is a second message here as well: schools that hope to identify those students most in need of advanced instruction in a particular domain should measure accomplishment in that domain, not in a distal or more general domain.

Long-term Predictions

In any domain, the best predictor of current performance is generally past performance on the same or similar tasks. Although the profile of students' reasoning abilities and other aptitudes can usefully inform *how* to teach students (Lohman & Hagen, 2001b), *what* to teach is best guided by what students know and can do. Therefore, short-term educational decisions should rely primarily on

Table 1

Percent of Students at Each Grade Who Scored above the 97th PR on ITBS Reading Total Identified When Other Selection Scores Are Used

Grade	ITBS		CogAT			
	Reading Total	Composite	Regression Estimate	Composite	Verbal	Nonverbal
3	100	51	38	38	36	19
4	100	57	36	31	34	22
5	100	56	36	29	36	15
6	100	52	36	29	35	17
Mean	100	54	36	32	35	18

Table 2

Percent of Students at Each Grade Who Scored above the 97th PR on ITBS Mathematics Total Identified When Other Selection Scores Are Used

Grade	ITBS		CogAT			
	Mathematics Total	Composite	Regression Estimate	Composite	Quantitative	Nonverbal
3	100	50	43	42	32	23
4	100	43	39	39	32	27
5	100	53	44	42	34	27
6	100	47	38	34	33	21
Mean	100	48	41	39	33	25

evidence of current accomplishment in a domain. Put differently, the primary “treatment” that educational institutions can offer is instruction commensurate with the students’ observed levels of achievement in particular domains. Immediate placement is best made on the basis of observed accomplishments in those domains.

Other aptitudes enter the picture, though, with each step one takes into the future. For example, given the same type of instruction, continued improvement in a domain requires interest or at least dogged persistence. More commonly, continued success requires a new mix of abilities: algebra requires skills not required in arithmetic; critical reading requires skills not required in beginning reading. Teachers, teaching methods, and classroom dynamics also change over time, each requiring, eliciting, or affording the use of a somewhat different set of person characteristics. Indeed, in most disciplines, the development of expertise requires mastery of new and, in some cases, qualitatively different tasks at different stages. Sometimes the critical factor is not only what is required for success but what is allowed or elicited by the new context that might create a stumbling block for the student. For example, in moving from a structured to a less structured environment, a student may flounder because he is anxious or is unable to schedule his time. Indeed, I sometimes think that the attainment of expertise has as much to do with inaptitudes as aptitudes.

The impact of these sometimes subtle changes in the demands and affordances of instructional environments is not obvious on summary measures of achievement to date. Scores on achievement tests show considerable year-to-year consistency. For example, the one-year stability of the Total Mathematics Score on the ITBS is about $r = .92$. However, even with this degree of stability there is much movement across several grades. In a longitudinal study of 6,321 Iowa students, the observed correlation between ITBS Mathematics Total scores at grade 3 and grade 8 was $r = .73$ in the top 3% of the mathematics distribution at grade 3 did *not* score in the top 3% of the distribution at grade 8. Prior achievement is thus not the only factor one must consider in predicting academic success over longer periods.

What are the other predictors of long-term academic success? In general, the second most important learner characteristic in the prediction of achievement is the ability to reason well in the symbol system(s) used to communicate new knowledge in a domain. Academic learning relies

heavily on reasoning (a) with words and the concepts they signify and (b) with quantitative symbols and the concepts they signify. Thus, the critical reasoning abilities for all students (minority and majority) are verbal and quantitative. Nonverbal (or figural) reasoning abilities are less important and show lower correlations with school achievement (Lohman, in press; Thorndike & Hagen, 1987, 1997).

Therefore, if the goal is to identify those students who are most likely to show high levels of future achievement, both current achievement and domain-specific reasoning abilities need to be considered. My analyses of the CogAT–ITBS data (Lohman, in press) suggest that the two should be weighted approximately equally. However, the relative importance of prior achievement and abstract reasoning depends on the demands and affordances of the instructional environment and on the age and experience of the learner. In general, prior achievement is more important when new learning is like the learning sampled on the achievement test. This is commonly the case when the interval between old and new learning is short. With longer time intervals between testings or when content changes abruptly (as from arithmetic to algebra), reasoning abilities become more important (Rock, Centra, & Linn, 1970). Novices typically rely more on knowledge-lean reasoning abilities than do domain experts. Because children are universal novices, reasoning abilities are more important in the identification of academic giftedness in children, whereas evidence of domain-specific accomplishments is relatively more important for adolescents.

Whether one is making short-term predictions about continued success in a particular educational context or long-term predictions about success in a new context, however, the critical issue is the identification of those aptitudes needed for success and of the inaptitudes that will thwart it.

The Prediction of Achievement for Minority Students

The selection policies used by some schools implicitly assume that the aptitude variables that best predict future academic success are different for minority than for majority students. For example, using a nonverbal test to identify academically gifted minority students presumes that nonverbal reasoning abilities are better indicants of academic aptitude for such students than measures of verbal or quantitative reasoning. Are the predictors of academic achievement the same for majority and minority students? For example, is the ability to reason with English words less predictive of achievement for

Hispanic or Asian-American students than for White students?

Elsewhere (Lohman, in press), I have reported analyses that address this question in some detail. Those analyses, which concur with those of other investigators (e.g., Keith, 1999), are unequivocal: The predictors of achievement in reading, mathematics, social studies, and science are the same for White, Black, Hispanic, and Asian-American students.

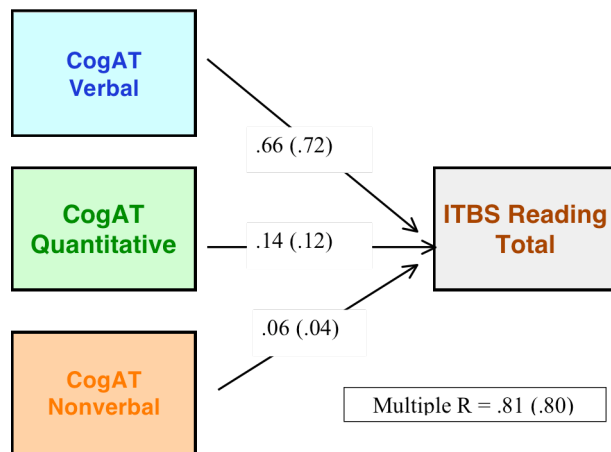


Figure 1. Average regression weights across grades 1 to 6 for the prediction of ITBS Reading Total scores from CogAT Verbal, Quantitative, and Nonverbal reasoning abilities. First weight is for non-Hispanic White students; the second weight (in parentheses) is for Hispanic students.

For example, Figure 1 shows how scores on the three CogAT batteries combine to predict ITBS reading achievement. Two regression weights are shown above each path. The first is for non-Hispanic White students; the second (in parentheses) is for Hispanic students. Clearly, the predictors of success in reading are the same for both groups. CogAT verbal reasoning is the strongest predictor; CogAT nonverbal reasoning contributes least to the prediction. Indeed, nonverbal reasoning abilities often have a negative regression weight in the prediction of achievement once verbal and quantitative reasoning abilities are in the equation (Lohman, in press; Case, 1977). This means that some students with high nonverbal reasoning scores are actually *less* likely to achieve well in school than other students with similar levels of verbal and quantitative abilities (see Lohman, in press).

This makes sense from the perspective of aptitude theory. Success in schooling places heavy demands on a student's abilities to use language to express her thoughts and to understand other people's

attempts to express their thoughts. Because of this, students most likely to succeed in formal schooling in any culture will be those who are best able to reason verbally. Indeed, our data show that, if anything, verbal reasoning abilities are even *more* important for bilingual students than for monolingual students.³ Thus, an aptitude perspective leads one to look for those students who have best developed the specific cognitive (and affective) aptitudes most required for acquiring expertise in particular domains. Identifying such students requires this attention to proximal, relevant aptitudes, not distal ones that have weaker psychological and statistical justification.

Assumptions about Growth

Judgments about aptitude invariably make assumptions about students' opportunities to learn the task from which inferences about aptitude are made. Inferences of aptitude from comparisons with grade peers presume that the pattern of a student's school attendance approximates that of other students in the same grade, that test and instructional content are aligned, and that out-of-school experiences which impact school achievement are similar. Comparisons with age peers presume that the student's general exposure to and participation in the culture sampled by the test approximates that of other students who are the same age. These assumptions are questionable for many students, and clearly false for some.

Predictions about future performance assume that the student's rank within group on the aptitude test will remain relatively constant over time. Note that this does not mean that one assumes that scores are fixed. Scores that report rank within age or grade group easily mask the fact that all abilities are developed; all respond to practice and instruction. Rather the assumption is that the student's rate of growth on the skills measured by the test will be the same as for others in the norm group who obtained the same initial score.⁴ This is unlikely either if the student's experiences to date differ from those in the norm group or if her subsequent experiences depart from the norm. For example, lack of experience in a domain will lead to a lower initial rank than the student will later achieve as she has the necessary learning experiences. This is especially true for

³ For an interpretation of the negative weight for nonverbal abilities, see Lohman (in press).

⁴ Depending on how the test is scaled, high-scoring students may need to gain more, the same, or less than low-scoring students in order to maintain their rank within group over time. In general, if the variance of scores increases over time, then they will need to gain more, and if it decreases, they will need to gain less.

well-defined skill sets (e.g., learning the letters of the alphabet) rather than for open-ended skill sets (e.g., verbal comprehension). But a student can also fall behind over time by improving, but at a slower rate than her peers. In general, prediction equations for academic success do not differ by ethnicity. Indeed, sometimes, aptitude tests overpredict the academic performance of some minority students (Willingham, Lewis, Morgan, & Ramsit, 1990). Thus, programs that aim to help minority students move from the high-potential to the high-accomplishment group might best understand their task as one of falsifying a prediction about growth rate.

This is not easily done. Contrary to popular myth, complex skills and deep conceptual knowledge do not suddenly emerge when the conditions that prevent or limit their growth are removed (cf. Humphreys, 1973). The attainment of academic excellence comes only after much practice and training. It requires the same level of commitment on the part of students, their families, and their schools as does the development of high levels of competence in athletics, music, or in other domains of nontrivial complexity.

The Fallacy of Common Cut Scores

Although the differences between minority and majority students are sometimes smaller on verbal and quantitative ability tests than on verbal and quantitative achievement tests, the differences are still substantial. A selection policy that uses either ability or achievement tests alone or that combines, say, mathematics achievement and quantitative reasoning ability will select proportionately fewer Black and Hispanic students than White and Asian-American students. How, then, can one attend to the relevant aptitude variables and increase the representation of underrepresented minority students?

Note that the discussion in this section concerns the identification of high-potential—not high-accomplishment—students. Current accomplishment, although perhaps measured in somewhat different ways for different individuals, should always be evaluated against the same high standards. That more White or Asian-American students achieve at high levels is problematic only if the selection tests are biased against other students. That this is not the case is widely accepted by measurement professionals (Jencks, 1998).

The identification of potential is a much slipperier task. Even in the best of circumstances, correlations between measures of aptitude and future achievement are lower; so predictions will often be wrong. More importantly, one can make inferences

about aptitude from a collection of tasks only when the individuals being compared have had similar opportunities to develop the skills required for success on those tasks. All recognize that many students—especially those whose first language is not English—have not had the same opportunities to develop skills in the English language. Therefore, when estimating the verbal reasoning abilities of such students, many look for tests that measure reasoning but that do not require facility with the English language. Unfortunately, there is no way to measure verbal reasoning skills without recourse to language! One can measure figural reasoning abilities that are correlated with verbal reasoning, but nonverbal reasoning abilities are as different from verbal reasoning as a test of physical fitness is from a test of basketball or ballet skills. And as with these psychomotor domains, the differences are most obvious at the extremes of the distribution. Furthermore, nonverbal reasoning tests do not identify the same students as tests of verbal or quantitative reasoning abilities (Lohman, in press). In other words, the assumption that all measures that load highly on G are exchangeable as selection tests is simply false. (See also Tables 1 and 2.)

Schools also use more distal aptitude tests because differences between English Language Learners and native speakers of English are sometimes smaller on such tests.⁵ The desire to use a common test with a common cut score for all applicants not only appeals to the laudable desire to be fair but also enormously simplifies the identification process. However, the consequences of such a policy far outweigh its benefits. Some of the more obvious deleterious effects are that it

1. *Reinforces the tendency to interpret intelligence and other ability tests as measuring innate abilities.* If scores on ability tests depend on background and education, then one must take these factors into account when interpreting them. The alternative—to interpret test scores as measures of innate abilities largely unaffected by such factors—avoids these complications. Thus, the decision to use a common cut score on aptitude tests inadvertently encourages the naïve but false belief that ability tests measure innate rather than developed abilities.

⁵ Differences are especially large when comparing nonverbal and verbal reasoning scores of ELL students. Differences are much smaller between quantitative and nonverbal reasoning tests, especially for Asian-American students. As a group, Black students often perform better on verbal and quantitative tests than on nonverbal reasoning tests (see, e.g., Jencks & Phillips, 1998).

2. *Encourages the use of less reliable tests.*

The smaller the mean difference between groups on the selection test, the greater the proportion of students from lower-scoring groups who will be selected using a common cut score. In general, group differences will be smaller on less reliable tests than on more reliable tests. For example, performance tests are generally less reliable than objective tests, and thus will generally show smaller group differences than objective tests. In the extreme, a completely unreliable test will show no differences between groups even when true differences are large. Therefore, evaluating tests by the extent to which they achieve the goal of proportional representation will tend to favor shorter and otherwise less reliable tests over longer and more reliable tests.

3. *Encourages the use of less valid tests.*

The hope that one can use a common cut score for all applicants leads one to opt for selection tests on which group differences are smaller. In general, though, when differences in achievement are large, differences will also be large on measures that predict achievement. Tests that are less predictive of achievement are more likely to show somewhat smaller group differences. For example, nonverbal ability tests show smaller differences between ELL and native speakers than verbal reasoning tests. However, such tests are also much poorer predictors of school achievement than verbal reasoning tests. Using less valid tests and a common cut score one may identify more minority students, but fewer who have the aptitude to succeed. This should be of concern to all, especially the minority communities who hope that the students who receive extra assistance will develop into the next generation of minority scholars and professionals.

A better policy, then, is to make decisions about potential for academic excellence using the most valid and reliable aptitude measures for all students and to compare each student's scores only to the scores of other students who share similar learning opportunities or background characteristics. In other words, identification of aptitude should be made within such groups. Those who balk at this suggestion might consider how commonly we shift among different norm groups when making evaluations about giftedness.

The Importance of the Norm Group

Grade cohort. Consider the second-grade child who scores at the 90th percentile rank (PR) in Reading Total on Form A of the ITBS. The student's

performance, while not exceptional, is certainly strong. But a norm group is implicit in this statement. Here, the norm group is students in the U.S. who were administered the test in approximately the same month of the 2000-2001 school year. Changing the norm group changes the percentile rank, sometimes subtly, sometimes substantially. For example, a November performance that rates a 90th PR using fall norms rates only a PR of 81 if midyear norms are used. In an effort to account for this ever-shifting achievement norm, test publishers typically use tables that estimate norms in *weekly* intervals. Clearly, though, interpretation of a given PR changes if one knows that the student missed several months of schooling due to illness or, less obviously, received more or less out-of-school instruction than other students on the skills sampled by the test.

Local norms. Although comparisons to the national norm group are useful for talent searches and other programs in which students will be grouped with students from other schools, the critical issue for most educational programming is the relative discrepancy between the student's performance and that of other students in the same instructional cohort. Indeed, students rarely find themselves in classrooms that represent the national distribution of abilities. For example, by midyear, the ITBS Reading Total Score that earned a 90th PR for individuals on Fall norms would actually be at the median in about 5% of classrooms in the nation. This means that in such classrooms, the student's Local Percentile Rank would be approximately 50. Conversely, in low-scoring school districts or classrooms, the same performance could easily fall above the 99th percentile. In short, although both national and local norms have important uses, decisions about acceleration are best made on the basis of local norms. These are offered by many test publishers when a school or district tests all children in a particular grade.

Age norms. Suppose, however, that we discover that the student whose achievement is exceptional is actually a year or more older than other children in the class. For example, some parents hold a child out of school for a year in order to give the child an advantage in physical and cognitive development over his or her classmates. Although instruction should be geared to the child's achievement, would one still consider the child "gifted"? Conversely, suppose a child is considerably younger than her classmates or has attended school irregularly. In both cases, comparisons with age peers can usefully inform judgments about academic giftedness. Tests that provide both age and grade norms allow comparison with both cohorts. This is useful when

the child is older or younger than grade peers. It is particularly helpful when the content of the test reflects general cognitive development rather than specific skills taught in school. Well-constructed ability tests provide this sort of information.

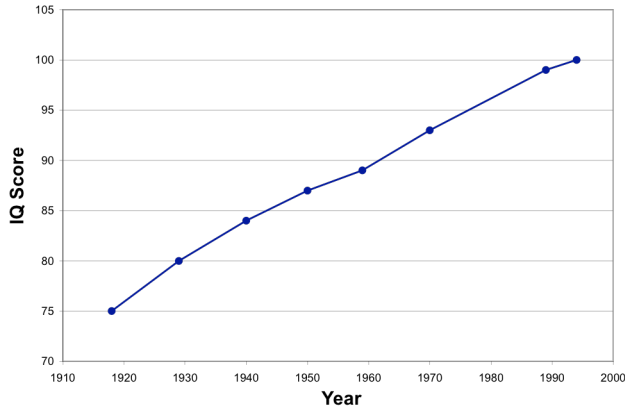


Figure 2. Estimated mean IQ scores for the Binet and Wechsler tests on the 1998 IQ scale by year in which the test was normed. (After J. Horgan, 1995, *Scientific American*, 273 (5), p. 14.)

Flynn effect. Norms for both ability and achievement tests change over time. The much-documented rise of scores on ability tests over the past seventy years (Flynn, 1999; Thorndike, 1975) makes it imperative that schools use tests with recent norms. Gains have been particularly large on figural reasoning tests such as the Raven Matrices. Broader measures such as the Stanford-Binet and Wechsler scales have shown smaller, but consistent, gains of about three IQ points per decade. Figure 2 shows one estimate of these changes. The examinee who obtained an IQ of 100 in 1998 would have received a score of 125 for a comparable performance in 1917.

Scaling effects. IQ scores are simply age percentile ranks reported on a different scale. An IQ of 100 always translates to an age PR of 50. The PR equivalent of other IQ scores depends on the standard deviation that is observed or assumed. For example, if $SD = 16$, then an IQ of 125 corresponds to an (age) PR of 94. If the SD is some other value, or if the distribution of scores is assumed to be positively skewed (rather than normally distributed), then a given PR may be associated with different IQ scores. For example, changes in the scaling of the Stanford-Binet between Form L-M and the fourth and fifth editions dramatically reduced the number of extremely high IQ scores that were reported (Ruf, 2003).

In short, judgments about exceptionality depend importantly on the norm group that is used. Whether a particular score is considered exceptional also depends on how the norms were derived, how the test scores were mapped onto a score scale, and how the scores will be interpreted. The child who is considered gifted when compared to others in his class may not be considered gifted when compared to others in the nation or to his age peers, to those who were tested a few months earlier or later, or to examinees of the same age who were tested a decade or two later.

In like manner, the score that indicates unusual verbal ability for a second-grade ELL student when compared with other ELL students may be unremarkable for the native speaker of English. The ELL student may have acquired English skills at a remarkably rapid rate when compared to other students with similar exposures to the English language. Although the student's current competence in using English when compared with others in the larger norm group may be well estimated by the test, inferences about her aptitude require a more focused comparison group.

However, test publishers do not report separate norms for different ethnic groups. There are many reasons for this, not the least of which are the difficulties that attend getting truly representative samples of different ethnic groups or the subsequent difficulties that would attend score interpretation. For example, achievement is generally best compared to a common set of standards. It makes little sense to set different standards for achievement when students must live and work in a common world. Nonetheless, inferences about aptitude that are sometimes made from test scores presume that examinees have had similar opportunities to acquire the knowledge and skills that are sampled by the test. I refer here not to the case in which inferences are made about innate ability, which are never justified, or inferences about current level of competence on the skills measured by the test, which generally are justified, but to inferences about ability to learn. The issue is particularly important when test scores are used to identify minority students who do not currently achieve at an exceptional level but who are most likely to develop academic excellence if given additional assistance. Such comparisons are best made by comparing a student's scores on the relevant aptitude test to those of other students who have had similar opportunities to develop the knowledge and skills measured by the test.

Even though many high-potential students identified in this way will not be ready for instruction at the same level as their high-accomplishment peers, are they ready for intensive instruction in advance of that received by their classmates? Suppose that we identified the top 3% of Black or Hispanic students and compared their scores to those of all other students. Where would they rank on the common scale? Following earlier analyses of reading and mathematics, we estimated aptitude for future achievement in each of these domains from students' observed achievement and the best prediction of their achievement from the three CogAT reasoning scores. We weighted observed and predicted achievement equally and then selected the top 3% of Black, Hispanic, and all students. Where did the best Black and Hispanic students fall on this common scale? In both reading and math, the typical Black student fell at the 90.8 PR in Reading and at the 91.5 PR in Math; the typical Hispanic student fell at the 93.9 PR in Reading and 94.8 PR in Math. Clearly, these are quite capable students. Change the norm group by comparing them to a slightly younger cohort of majority students or to students of an earlier generation, and all would be considered "gifted"—at least on this measure of learning potential. Nonetheless, many of these students are achieving at levels well below those whose achievement scores alone place them at the top of the group. This means that high-potential students may have different instructional needs than high-accomplishment students, especially in hierarchically ordered domains such as mathematics.

Suggestions for Policy

How could a school implement a policy that would be consistent with the principles outlined here? Consider the following policy points:

1. *What educational treatment options are available?* Understanding the treatment is the first step in understanding what person characteristics will function as aptitudes (or inaptitudes) for those treatments. Will students receive accelerated instruction with age-mates, or will they be grouped with older children whose achievement is at approximately the same level? Will instruction require much independent learning or must the student work with other students? Will instruction build on students' interests or is the curriculum decided in advance? These different instructional arrangements will require somewhat different cognitive, affective, and conative aptitudes. At the very least, different instructional paths should be available for those who already exhibit high

accomplishment and those who display potential for accomplishment. For those in the former group, acceleration or, if you wish, "developmentally appropriate instructional placement," is often the most effective treatment. For those in the latter group, special programs that provide intensive instruction designed to develop competence are needed. If schools cannot provide this sort of differential placement, then it is unlikely that they will be able to satisfy the twin goals of providing developmentally appropriate instruction for academically advanced students while substantially increasing the number of underrepresented minority students who are served and who subsequently develop academic excellence.

2. *Decide the extent to which selection is to be based on evidence of accomplishment or on potential for accomplishment.* In general, emphasize accomplishment when identifying academically gifted older children and adolescents. Emphasize potential for young children and for those who have not had the opportunity to attain significant levels of expertise in a domain. However, at all ages, evidence of high current accomplishment should trump predictions about future accomplishment, especially when deciding what to teach.

3. *Establish policies for achieving more equitable representation of minority students in programs.* Discuss the difference between the need for common standards in the measurement of current achievement and the need for within-group standards for the measurement of potential. Setting common, high standards for all encourages those who do not yet display these skills to work toward them. Because the discrepancy between potential and accomplishment will be greatest for those who have had the fewest opportunities, consider weighting accomplishment more heavily for advantaged students and potential for students whose educational opportunities have been more limited. Or keep the weights the same for all but group students by opportunity to learn and make selections within groups. Then make instructional placements primarily on the basis of accomplishments to date. If procedures like these were used to identify Black and Hispanic students, schools could have much greater confidence that they had identified the most academically promising minority students. Common cut scores on less valid and reliable selection tests may identify significant numbers

of minority students, but many of them will not succeed in an advanced program.

Keep in mind that there is also an ethical dimension to be considered. For some children, the intensive instruction offered in special programs for the gifted provides opportunities that supplement what their families provide; for other children, the same programs provide the only opportunity to develop academic skills. Indeed, the goal for these students is to provide educational opportunities that will falsify the prediction that future achievement will show the same or lower rank than current achievement.

4. *Obtain the most reliable and valid measures of proximal achievement and aptitude variables for all students.* Do not base selection on composite scores on achievement or ability, especially for older students. Rather, obtain measures of domain-specific achievement, the student's ability to reason in the symbol systems required for new learning in that domain, interest in the domain, and persistence under similar instructional conditions. For example, to identify students who currently excel in mathematics, measure mathematics achievement using a well-constructed, norm-referenced achievement test that emphasizes problem solving and concepts rather than computation. Consider using an out-of-level test if the student may be accelerated to a higher grade. To identify students who currently do not exhibit superior mathematical competence but who show potential to develop it, combine scores on the mathematics achievement test with scores on a well-constructed, norm-referenced measure of quantitative reasoning ability. Generally, combine the scores in a way that weighs mathematics achievement and quantitative reasoning abilities equally. To assess interests, inquire specifically about the students' interests in mathematics or in occupations that require mathematical thinking. Interest inventories can be helpful, especially for adolescents (see Lubinski, Benbow, & Ryan, 1995). Finally, persistence is best estimated from ratings of persistence by teachers and others who have worked with the child in situations like those to be encountered in the planned acceleration program.

5. *Make better use of local norms when identifying students whose accomplishments in particular academic domains are well above those of their classmates.* For example, on norm-referenced achievement tests, look at local

percentile ranks for particular domains such as mathematics or science rather than at national percentile ranks for composite scores. Provide instruction that is developmentally appropriate, for example through acceleration. When you place students in another grade for instruction, consider out-of-level testing for measuring the students' academic accomplishments relative to their prospective peer group. For example, if students will be placed with seventh graders for mathematics, compare their mathematics achievement to seventh graders on a test with seventh-grade content. Although measuring achievement within domains will increase the representation of ELL students in mathematics programs, expect that the students selected will be disproportionately White and Asian American.

6. *Emphasize that true academic giftedness is evidenced by accomplishment.* Predictions that one might someday exhibit excellence in a domain are flattering but unhelpful if they do not translate into purposeful striving toward the goal of academic excellence. Indeed, the attainment of academic excellence requires the same level of commitment on the part of students, their families, and their schools as does the development of high levels of competence in any other domain. Students may find it helpful to consider identification as a "high-potential" student as analogous to being identified as a "high-potential" athlete and then to investigate the duration and intensity of training that high-caliber athletes endure in order to rise to the top of their sport. This also means that students must be identified with an eye on the kind of intensive instruction that can be offered. If advanced instruction will be in writing short stories, then measures of quantitative or figural reasoning abilities will not identify many of those who are most likely to succeed. Further, if possible, the instruction that is offered should be adapted better to meet the needs of minority students in developing the academic and personal skills that they will need to succeed in schooling. On the affective side, eliciting interest and persistence are critical. On the cognitive side, oral language skills are probably the most neglected, but among the most important. Many suggestions can be derived from case studies of successful minority scholars or from evaluations of schools that routinely produce them (e.g., Presseley, Raphael, Gallagher, & DiBella, 2004).

In any case, the concept of aptitude—although much maligned and even more commonly misunderstood—is critical in the identification process.

References

- Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist, 38*, 85-94.
- Assouline, S. G. (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.
- Bingham, W. V. (1937). *Aptitudes and aptitude testing*. New York: Harper & Brother.
- Carroll, J. B. (1974). The aptitude-achievement distinction: The case of foreign language aptitude and proficiency. In D. R. Green (Ed.), *The aptitude-achievement distinction* (pp. 286-303). Monterey, CA: CTB/McGraw-Hill.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.
- Case, M. E. (1977). *A validation study of the Nonverbal Battery of the Cognitive Abilities Test at grades 3, 4, and 6*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. E. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54*, 5-20.
- Gustafsson, J. -E., & Baulke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*, 407-434.
- Hagen, E. P. (1980). *Identification of the gifted*. New York: Teachers College Press.
- Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Tests of Basic Skills: Form A*. Itasca, IL: Riverside.
- Horgan, J. (1995). Get smart, take a test. *Scientific American, 273* (5), 14.
- Humphreys, L. G. (1973). Implications of group differences for test interpretation. In *Proceedings of the 1972 Invitational Conference on Testing Problems: Assessment in a Pluralistic Society* (pp. ____). Princeton, NJ: ETS.
- Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 55-85). Washington, DC: Brookings Institution Press.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly, 14*, 239-262.
- Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 3*, 482-494.
- Lohman, D. F. (2004). *Tables of prediction efficiencies*. Retrieved August 21, 2004, from The University of Iowa, College of Education Web site: <http://faculty.education.uiowa.edu/dlohman/>
- Lohman, D. F. (in press). The role of nonverbal ability tests in identifying students for participation in programs for the academically gifted. *Gifted Child Quarterly*.
- Lohman, D. F., & Hagen, E. P. (2001a). *Cognitive Abilities Test (Form 6)*. Itasca, IL: Riverside.
- Lohman, D. F., & Hagen, E. P. (2001b). *Cognitive Abilities Test (Form 6): Interpretive guide for teachers and counselors*. Itasca, IL: Riverside.
- Lohman, D. F., & Hagen, E. P. (2002). *Cognitive Abilities Test (Form 6): Research handbook*. Itasca, IL: Riverside.
- Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist, 55*, 137-150.
- Lubinski, D., & Benbow, C. P., & Ryan, J. (1995). The stability of vocational interests among the intellectually gifted from adolescence to adulthood: A 15-year longitudinal study. *Journal of Applied Psychology, 80*, 196-200.
- Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24*, 159-201.

- Martin, D. J. (1985). *The measurement of growth in educational achievement*. Unpublished doctoral dissertation, The University of Iowa, Iowa City.
- McGrew, K. S., & Evans, J. J. (2004). *Internal and external factorial extensions to the Cattell-Horn-Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise* (Carroll Human Cognitive Abilities Project Research Report No. 2). Retrieved July 20, 2004, from the Institute for Applied Psychometrics Web site: <http://www.iapsych.com/carrollproject.htm>
- Peterson, P. L. (1977). Interactive effects of student anxiety, achievement orientation, and teacher behavior on student achievement and attitude. *Journal of Educational Psychology, 69*, 779-792.
- Presseley, M., Raphael, L., Gallagher, J. D., & DiBella, J. (2004). Providence-St. Mel School: How a school that works for African American students works. *Journal of Educational Psychology, 96*, 216-235.
- Rock, D. A., Centra, J. A., & Linn, R. L. (1970). Relationships between college characteristics and student achievement. *American Educational Research Journal, 7*, 109-121.
- Ruf, D. L. (2003). Use of the SB5 in the assessment of high abilities (Stanford-Binet Intelligence Scales, Fifth Edition: Assessment Service Bulletin No. 3). Itasca, IL: Riverside.
- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*, 604-614.
- Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*, 347-376.
- Snow, R. E., & Yalow, E. (1982). Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493-585). Cambridge, MA: Cambridge University Press.
- Subotnick, R., & Jarvin, L. (in press). Beyond expertise: Conceptions of giftedness as great performance. In R. J. Sternberg & J. Davidson (Eds.), *Conceptions of giftedness* (2nd ed.). New York: Cambridge University Press.
- Tanner, J. M. (1965). Physique and athletic performance: A study of Olympic athletes. In S. A. Barnett & A. McLaren (Eds.), *Penguin science survey 1965 B* (pp. 112-133). Baltimore: Penguin Books.
- Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4*, 3-7.
- Thorndike, R. L., & Hagen, E. P. (1987). *Cognitive Abilities Test (Form 4): Technical manual*. Chicago: Riverside.
- Thorndike, R. L., & Hagen, E. P. (1997). *Cognitive Abilities Test (Form 5): Research handbook*. Chicago: Riverside.
- Traub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll Theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19*, 72-87.
- Willingham, W. W., Lewis, C., Morgan, R., & Ramsit, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. New York: The College Board.