Aptitude for college: The importance of reasoning tests for minority admissions

David F. Lohman

The University of Iowa

Draft of October 3, 2002

Revision of comments on Dr. Atkinson's proposal given at

Rethinking the SAT*: The future of standardized testing in university admissions,*

University of California, Santa Barbara, November 16, 2001.

College admissions tests are expected to serve multiple and often contradictory purposes. Because of this fact, an admissions test that serves one purpose well may serve other purposes poorly. The two most important purposes of admissions tests are (a) to report on students' academic development to date and (b) to predict the likelihood of their success in college. Each of these goals is hard to accomplish with one test; achieving both with the same test may be impossible. Because of this, I argue that aptitude tests that go beyond prior achievement have an important role to play in admissions decisions, especially for minority students.

Before embarking on a discussion of aptitude testing, it is helpful to consider briefly a few of the difficulties that attend the seemingly simple goal of reporting on the level of academic knowledge and skill that students have acquired during their previous schooling. Achievement tests that are closely aligned to the common curriculum are most useful for this purpose. Such tests can help focus the efforts of both teachers and students on the knowledge and skills that will be used to make admissions decisions. This is generally viewed as a good thing unless, as seems often the case with high-stakes tests, the test unduly narrows the curriculum. Furthermore, high school students – especially those in different regions of the country -- experience different curricula, so a test that represents the common curriculum must focus on students' general educational development. Such tests often contain relatively little specific content or conceptual knowledge. Science tests, for example, typically include tasks that require examinees to show that they can set up experiments or engage in other forms of scientific reasoning; but they generally do not sample what students might know about the periodic table or the function of the respiratory system. Again, some view this as a good thing. They believe that science should be about process, not content. Others think that ignoring content knowledge distorts the measurement of what students have learned, especially poor children who attend schools that

emphasize learning content and basic skills more than problem solving and critical thinking. Finally, some argue that achievement tests should present tasks that mimic authentic performances, such as conducting science experiments, writing essays on topics of personal interest, or reasoning mathematically about ill-structured problems. Others argue that this is not always possible or desirable. In short, the seemingly simple goal of reporting on what students know and can do is not as straightforward as it might seem.

The second purpose of admissions tests is to look forward and predict the likelihood of a student's success in some yet-to-be-experienced environment. This aspect of admissions testing is less clearly represented in the current debate. The key concept here is aptitude, specifically aptitude for academic learning in different university programs. Dr. Atkinson is right when he complains about "ill-defined notions of aptitude." But the concept of aptitude—whether well or poorly defined—is central to this discussion.

Aptitude

Students arrive at the university with characteristics developed through life experiences to date. These include their knowledge and skills in different academic domains, their ability to think about fresh problems, their motivation and persistence, their attitudes and values, their anxiety levels, and so on. The university experience may be conceptualized as a 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. Specifically, aptitude refers to *the degree of readiness to learn and to perform well in a particular situation or fixed 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. Examples include the ability to take good notes, to manage one's time, to use previously acquired knowledge appropriately, to make good inferences and generalizations, and to manage one's emotions. Aptitudes for learning thus go beyond cognitive abilities. Aspects of personality and motivation commonly function as aptitudes as well.

However, the same situation can evoke quite different ways of responding in different individuals. As a result, different measures of aptitude may be required to predict the performance of students who follow different routes to academic success. Because of this, a good selection system must cast a broad, not a narrow net. It must also look carefully at the demands and opportunities of different university environments, since defining the situation is part of defining the aptitude.

An example may help clarify how the same situation can evoke predictably different ways of responding. Students who come from different segments of our society often find the same college environment to be more or less congenial with their prior school experiences. For some, the situation will be familiar and will allow the use of practiced ways of responding. For others, however, the same situation will require more personal adaptation and novel problem solving. One of the factors that moderate such relationships is social class. Educational reformers are once again rediscovering the extent to which reforms that emphasize independent thinking in mathematics, for example, are often better received by middle- and upper middle-class students than by lower-class students (e.g., Lubienski, 2000). If the goal is to find lower-class students who are likely to succeed in college and beyond, then one must sample more than the curriculum that committees of middle-class educational reformers prefer and that middle-class students are

likely to have experienced. Put differently, one must have a view of aptitude that embraces more than past achievement of the "common" curriculum.

One possibility is to use test-tasks that sample abilities students have developed through their everyday experiences. Given the diversity of such experiences, one must find a way quickly to sample the sophistication of the students' reasoning in a broad range of contexts. Those who study reasoning abilities have investigated many different ways of constructing items to do this. Analogies repeatedly emerge as one of the most efficient item types. Although the format is ancient, research on how people solve such problems is extensive and recent. Dr. Atkinson rightly argues that verbal analogy items should not be the object of instruction, and that some analogy items seem primarily to test vocabulary knowledge. But eliminating such items will not necessarily produce a better test. The analogy format allows one to sample the efficacy of both past and present verbal reasoning processes across a much broader range of domains than could ever be represented in a necessarily smaller sample of reading passages. And even though good analogy items require more than vocabulary knowledge, word knowledge is not as irrelevant as it might seem to be. Indeed, well-constructed vocabulary tests are among the best measures of verbal reasoning. This is because students learn most new words by inferring their meanings from the contexts in which the words are embedded, and then remembering and revising their understandings as they encounter the word anew. Achieving precise understandings of relatively common, but abstract words is thus an excellent measure of the efficacy of past reasoning processes in many hundreds or thousands of contexts. On the other hand, knowledge of infrequent or specialized words, while sometimes useful as a measure of prior achievement, estimates reasoning poorly and thus should be avoided on an aptitude test that aims to measure reasoning rather than domain knowledge.

However, public debates about testing policies rarely deal in such subtleties (Cronbach, 1975). Appearances matter more than substance, so if analogy items appear inauthentic or problematic they will be (indeed, now have been!) eliminated. The sad part of this story is that many of those most enthused by this change will benefit least from it. For its part, ETS has not always built analogy items in ways that would allow reasonable defense of the reasoning construct they are intended to measure. Indeed, it is possible to build good analogy items for 12th graders using words that most 7th graders know.[1]

A revisionist history of the SAT

The untold story of the SAT is really about how the concept of aptitude was at first embraced, then simply assumed, then became an embarrassment, and, most recently, abandoned. The problem with a word such as "aptitude" is that everyone thinks that they know what the word means, so they are not inclined to check their understandings against more careful expositions. This is a common problem in psychology. Many key psychological constructs – such as learning, motivation, or intelligence – have deeply entrenched everyday meanings. Because of this, some psychologists have invented new terms for psychological constructs (e.g., Cattell, 1965) or have tried to abandon value-laden terms in favor of less value-laden terms. Jencks (1998) believes that the only way to eliminate this sort of "labeling bias" in ability tests is to re-label the test. This was the solution initially proposed by these who attempted (unsuccessfully, as it turned out) to change the middle name of the SAT from "aptitude" to "assessment." Unfortunately, there is no value-free synonym for "aptitude."

[1] For example, consider the verbal analogy items on the 12th grade level of the Cognitive Abilities Test (Lohman & Hagen, 2001). The typical correct answer is a word that can be used correctly in a sentence by about 75% of 7th graders. The average vocabulary level of all other words in the analogy items is grade 5. Nevertheless, the analogy items are quite difficult. The typical 12th grade student answers only about half of the items correctly.

The root of the problem is that Carl Brigham adopted the word "aptitude" in his test without a good theory of what aptitude might be. Brigham's background was in intelligence testing, so he (and many others) assumed that the intelligence tested by his test was the most important scholastic aptitude. Clearly, testing aptitude was Binet's original intent. He sought to devise a series of tests that would identify those who were unlikely to benefit from formal schooling and who would instead need special training. Harvard's President Conant also wanted to measure aptitude, but for the opposite purpose. His goal was to find students who were likely to succeed at Harvard but who had not attended one the handful of private schools from which Harvard selected most of its students. Why not use an achievement test instead? As Lemann (1999) observed "What Conant didn't like about achievement tests was that they favored rich boys whose parents could buy them top-flight instruction." (p. 38) Those who would rely solely on achievement tests to forecast college success still need to worry about this issue.

The history of the SAT might have been quite different had its founder been Walter Bingham instead of Carl Brigham. Whereas Brigham's background was in intelligence testing, Bingham's expertise was in what we would call industrial psychology. Bingham's (1937) Aptitudes and aptitude testing is still worth reading, especially in conjunction with some of the more tortured treatises on the aptitude-achievement distinction of later theorists who had greater competence in multivariate statistics than psychology. Industrial psychologists – from Clark Hull to Richard Snow – have always had more success in thinking about what aptitude might be than many of their counterparts in education. Predicting how well applicants are likely to succeed in a job for which they have little or no prior training is as commonplace in industry as it is uncommon in education. But it is when the mismatch between prior experience and future job demands is greatest that we must think most clearly about why some succeed while others fail.

Put differently, educators are easily lulled in to thinking that they understand *why* some succeed when at best they understand *who* has succeeded in the past. As long as both the system and the participants remain the same, those who succeeded in the past will indeed be the most likely to succeed in the future. But change either the individual or the system and the prediction fails.

The goal of aptitude testing, then, is to make predictions about the individual's likelihood of success and satisfaction in some yet to be experienced situation on the basis of present behavior. Bingham (1937) spoke of aptitude as readiness to learn some knowledge, skill, or set of responses. This "readiness to acquire proficiency" also included affective factors, such as interest or motivation. Bingham also emphasized that aptitude does not refer to native endowment but rather simply to present characteristics that are indicative of future accomplishment.

> Whether (a person) was born that way, or acquired certain enduring characteristics in his early infancy, or matured under circumstances which have radically altered his original capacities is … of little practical moment . . . And so, when appraising his aptitude, whether for leadership, for selling, for research, or for artistic design, we must take [the person] as he is – not as he might have been. (p. 17)

Unfortunately, this view of *aptitude* was less intuitively appealing than one that emphasized the contributions of biology. Early studies of the mental abilities of twins seemed to support beliefs that intelligence and other scholastic aptitude tests really did measure something innate (see, e.g., Lohman, 1997, for one explanation). The developers of the SAT had a more nuanced understanding, generally acknowledging that the abilities measured by the SAT were not innate and developed over time. Further, these abilities were said to be "influenced by

experience both in and out of school." (Donlon & Burton, 1984; p 125). But without a clear theory of what aptitude might be, such caveats were easily ignored.

In the educational literature, some of best early thinking about aptitude can be found in Carroll's writings about foreign language aptitude. Once again, this case is closer to the task faced by industrial psychologists than by those would predict the ability to read critically in college from similar reading abilities displayed in high school. In devising tasks for his foreign language aptitude test, Carroll could not assume prior proficiency in the foreign language. And so he sought to create test tasks that had a "process structure similar to, or even identical with, the process structures exemplified in the actual learning tasks, even though the contents might be different." (Carroll, 1974, p. 294). One cannot accomplish this goal unless one first has a reasonably good understanding of the processing demands of tasks in the target domain. Prior to the advent of cognitive psychology, understanding cognitive processes was largely a matter of having good intuitions. But we now know quite a bit about the cognitive demands of different instructional environments, and of the person characteristics that are necessary for and therefore predictive of success in those environments (Corno et al., 2002). In other words, we are in much better position to build aptitude tests today than Carl Brigham was back in the 1920's when he assembled the first edition of the SAT.

Aptitude testing, then, is not about measuring innate capacities – whatever these might be. Rather, it begins with a careful examination of the demands and affordances of the target environment and then attempts to determine the person characteristics that facilitate or impede performance in those environments. 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. Thus, the first task in developing a good

aptitude test is careful study of the target domain, especially of its demands and affordances. We need much more of this work at the university level. The second task is to identify those characteristics that predispose individuals to succeed in the environment. Prior knowledge and skill are often the single most important predictors of success in academic environments. But these are not the only person characteristics that matter. The ability to reason well in the symbol system(s) used to communicate new knowledge is particularly important for those who cannot rely as readily on well-developed systems of knowledge in the domain. Likewise, the ability to persist in one's efforts to attain a difficult goal is also critical for those who start the race several steps behind. This means that although achievement tests may better direct the efforts of students in secondary school, and report on the extent to which they have achieved the common curriculum, tests that measure reasoning abilities and other aptitudes for success in college can help admissions officers find students who are likely to succeed in spite of less than stellar performance on the achievement test. This leads to the next point.

Fluid-crystallized ability continuum

When discussing a selection system, it is helpful to keep track of the commonalities and differences among the measures that are used. One way is to track the extent to which different tests estimate students' abilities to solve familiar problems using practiced routines versus their abilities to solve unfamiliar problems using general reasoning abilities. Figure 1 shows such a continuum.

Insert Figure 1 here

Assessments differ in the extent to which they are tied to context and situation. For example, course grades are based on tests, projects, and other assignments that are tightly bound

to the particular learning context. Some psychologists refer to the knowledge and skill measured by such assessments as *crystallized* abilities. Averages of grades across courses are less tied to any one context. Achievement tests that aim to measure students' understanding of a common curriculum require more transfer. When there is no common curriculum or when we choose tasks that are deliberately novel for our assessments, then we move tasks even further to the left.

In other words, as we move from right to left on this continuum, we move from measures that are embedded in the curriculum to measures that have no obvious connection to the curriculum. The latter are sometimes called *fluid* reasoning abilities. To the extent that assessments are meant to inform students what they should know, tests near the right are clearly more useful. But to the extent that we want measures that have added value beyond high school grades, then we need to measure abilities at different points along this continuum. This is because assessments that are nearer each other will generally be more highly correlated. Students identified as likely to succeed in college by one test will tend to be the same students identified by the other test. In this regard, you will notice that although I have placed *SAT I* to the left of *SAT II* general tests and the *ACT* assessment, I have not placed them very far apart.

One way to think about the current debate is in terms of where along this sort of continuum college entrance tests should lie. Some favor moving towards the right. They do this in part because they want measures more closely aligned with the curriculum. Some do this because they treat freshman grade-point averages as the gold standard and seem not to realize that grades are only one of many possible measures of success in learning. Many also want to get as far away as they can from measures of reasoning abilities that remind them of IQ tests. Disdain for item types such as analogies is grounded in a legitimate concern that such item types should not be the object of instruction, in a legitimate concern for the extent to which knowledge

is indeed situated, but also in a failure to appreciate what we have learned about the measurement of human cognitive abilities in the past twenty years. Which leads to the next point.

The importance of fluid reasoning abilities in a selection system

It is commonly believed that tests of general reasoning abilities that use the sort of items once used on IQ tests are inherently biased against minorities. Some of these tests and some of the items on them were bad by any standard. But we have learned a thing or two about how people think and about how to measure thinking. *In fact, scores on well-constructed measures of developed reasoning abilities actually show* ***smaller*** *differences between white and minority students than do scores on good achievement tests*. And this is one of the main reasons why tests that measure reasoning abilities using non-academic tasks can be helpful in the admissions process. They can assist in identifying students who do not do particularly well on the more curriculum-based tests, but who are likely to succeed if they work hard.

Figure 2 shows data for of 11th grade students who participated in the joint 2000 National Standardization of the *Iowa Tests of Educational Development* (*ITED*) and the *Cognitive Abilities Test* (*CogAT*). The *ITED* is a general achievement test for high school students. Parts of the test are very similar to the *ACT* assessment; parts are similar to the *SAT*. It shows high correlations with both. The ITED score used here is the Core Total, without math computation. This total score includes tests for critical reading of literary materials, social studies, and science; reading vocabulary; correctness and appropriateness of expression; and mathematical concepts, problems, and interpretations. The *CogAT* measures reasoning abilities in three domains or symbol systems: verbal, quantitative, and figural (or nonverbal). The Nonverbal Battery is least tied to the curriculum. The item formats are ancient. They include

sentence completions, series completions, classification problems, matrix problems, and yes, even verbal and figural analogies. Although these item formats are old, the construction of items was informed by thirty years of research in cognitive psychology on how people solve such problems and how test items can be constructed better to measure reasoning abilities (see, e.g., Lohman, 2000).

Insert Figure 2 here

The question is: What is the percent of minority students who score above the 70th percentile on the *ITED* and each of the three batteries of the *CogAT*? Grade 11 was chosen because the data are generally more dependable than at grade 12, although here it makes little difference. The 70th percentile was chosen to insure a sufficient sample size for all four groups. Similar patterns are observed at higher cut points. Each column in Figure 2 scores shows the increment (or decrement) in percent of students who would be selected using a particular *CogAT* score versus the number who would be selected using the achievement test. Thus, the first bar in each set is fixed at 1.0. Look first at the data for White students. It makes little difference which test is used. Now look at the data for Blacks. All three reasoning tests—but especially the CogAT Battery—show increases over the achievement test in the percent of Black students who would be selected. For Hispanic students, the Battery shows a drop. Making nuanced judgments about the meanings of words in the English language is not a strength. However, Quantitative and especially Nonverbal reasoning scores are higher. Finally, for Asian Americans, the Quantitative and Nonverbal batteries are once again more likely to identify able students. Those who are concerned about the number of minority students admitted should be concerned about the kind of tests that are administered. The recent report by Bridgeman, Burton, and Cline (2001)

comparing the percent of minority students admitted under *SAT I* and *SAT II* did not find this difference. This reinforces my assertion that these tests are actually closer to each other than some would expect. Indeed, the problem with the current version of the *SAT I* may not be that it is an aptitude test, but that it is not enough of an aptitude test. Over the years it has become more and more achievement-like. The pressure to eliminate discrete item types (such as analogies and sentence completions) and include more "authentic" tasks promises to make the SAT even more like an achievement test. This means that there is a growing need for an alternative measure of students' abilities that is not so strongly tied to the goals of the common curriculum. Such a test could be offered as an optional but separate battery. It could provide important information for admissions committees when they are confronted with applications from poor and minority students who have not scored sufficiently well on the achievement-oriented tests, especially those who show evidence of persistence and a desire to succeed in school.[2]

It is important to understand that the differences between the CogAT and the ITED shown in Figure 2 are not due to bias in the achievement test. Much of the discussion about origin of social class differences in mental test scores (e.g., Eells, 1951) and the reaction of conservative psychometricians to it (e.g., Jensen, 1980 – who was a student of Eells) is based implicitly or explicitly on the assumption that a good test of mental ability should somehow be able to see through the veneer of culture and education to the "real" or "innate" differences that lie below the surface. That students who have had a superior education are better able to understand and critically examine the sort of abstract educational ideas presented on the ITED is no more surprising than the fact that those who have had better training in, say, basketball can

[2] For various reasons, I do not think that the test should be as removed from the curriculum as the Analytic subtest on the current GRE. Academic learning depends most heavily on student's abilities to reason with words and with quantitative concepts. But one can measure these abilities in ways that reduces the impact of formal schooling (see footnote 1).

participate at higher levels in that sport. Good measures of school learning must emphasize those aspects of the curriculum that educators value most. Nevertheless, good measures of reasoning abilities can be built that reduce the direct influences of schooling.

A related confusion is the expectation that measures of fluid reasoning abilities should better predict criteria such as course grades than do achievement test scores or grades in previous courses. In chemistry there is a saying "like dissolves like." In psychometrics of prediction, the parallel dictum is "like best predicts like." When freshman GPA is the criterion, then, other things being equal, high school GPA will generally be the best predictor, measures of achievement the next best predictor, and measures of fluid reasoning the weakest predictor. If common exams at the end of the first year of college were the criterion, then similar measures of past achievement would probably be the best predictor. And if the ability to solve unfamiliar problems in- or outside of one's field of study were the criterion, then measures of fluid reasoning in the same symbol system might top the list of predictors.

There is a large literature on the characteristics of persons and tasks that predict school learning. Correlations between college entrance tests and freshman grade-point average are a small and exceedingly unimportant part of that literature. Indeed, I am astonished that such studies show anything, given the diversity of grading criteria and course content across instructors, domains, and schools (see Young, 1990, for one effort to accommodate some of these factors). In their summary of the predictive validity of the SAT, Willingham, Lewis, Morgan, and Ramist (1990) conclude that "a simple analysis of the relationship between [freshman GPA] and scores on pre-admission predictors conceals almost as much as it reveals." (p. 84). Most notably, when the criteria are grades in particular courses rather than GPA, the SAT is a consistently better predictor than high-school GPA. In large measure, this is due to the

diversity of grading standards across courses that enter into the first-year GPA. Further, grades in large undergraduate classes are commonly determined by performance on objective tests. This means that the course grade may be simply a rough surrogate for two or three course-specific achievement tests. But there are many other ways to measure success in learning, and correlations among these measures typically show considerable divergence. Therefore, decisions about which students to admit should make a serious effort to gather and find the predictors of measures of academic success other than GPA. Continuing to accumulate information on the predictors of first-year GPA may help track local variation in this rather modest relationship, but little else. Looking at a diversity of learning outcomes within large classes can show the value of other measures. However, finding measures that best predict success in a given system can have the paradoxical effect of identifying those likely to succeed in a system that might be in dire need of repair. Indeed, one of the more important uses of measures of fluid and crystallized abilities in research on academic learning has been to find those instructional methods that *reduce* the relationship between learning success and reasoning abilities or prior achievement. Systematic declines in the predictive validity of both the SAT and high school GPA from 1970 to 1988 at some institutions may reflect such adaptations.

An analogy to physical skills

One should not infer that fluid reasoning abilities are innate and that crystallized achievements developed. Both fluid and crystallized abilities are developed. The primary difference lies in the extent to which abilities are developed through explicit, focused training and practice or are instead the more indirect outcomes of such experiences. But this is difficult to understand because our intuitive theories of abilities constantly get in the way. The best way I have found to understand the difference between ability (as aptitude) and achievement (as

outcome) is by analogy to physical skills. Let me return to the continuum of transfer shown in Figure 1. This time, however, the domain is physical skills rather than cognitive abilities (see Figure 3).

Insert Figure 3 here

Crystallized abilities are like knowledge and skill in playing different sports. These skills are developed through years of practice and training. Athletes show different levels of competence across sports just as students show different levels of competence in various school subjects. But athletes also differ in their levels of physical fitness. Physical fitness is aptitude for acquiring skill in most sports. Athletes who have higher levels of physical fitness or conditioning will generally have an easier time learning new skills and will perform those that they do learn at a higher level. But physical fitness is also an *outcome* of participation in physically demanding activities. Further some sports—such as swimming —are more physically demanding than other sports and result in higher increments in physical conditioning for those who participate in them. In a similar manner, reasoning abilities are both an input to as well as an outcome of good schooling (Snow, 1996; Martinez, 2000). Indeed, expecting a measure of reasoning abilities to be independent of education, experience, and culture is like expecting a measure physical fitness to be uninfluenced by the sports and physical activities in which a person has participated.[3]

The task of selecting students for university training is akin to selecting students who are likely to succeed in college-level athletics. The best predictor of students' abilities to play football or basketball in college is their demonstrated abilities to play those same sports in high school. In

[3] This analogy also acknowledges the importance of biological factors, but makes clear the absurdity of the all-too-common inference that an unbiased ability test (or unbiased test of physical fitness) should somehow not be influenced by experience.

like manner, the best indicator of their abilities to get good grades in college is their abilities to get good grades in similar courses in high school. When athletes come from small schools, however, evaluating athletic skill is difficult unless coaches can arrange a common competition such as a summer basketball tournament. Similarly, achievement tests in particular domains can provide a common yardstick across schools. Suppose, however, that when we have assembled our football team we are short of wide receivers, or on our basketball team, of someone to play center. The question, then, becomes one of finding people who are likely to succeed even though their athletic performance thus far has not been stellar. What we look for are athletes who have the requisite physical skills (such as strength, speed, or agility) and at least a moderate level of skill in the sport. Our intention would be not simply to put these athletes on our team but first to provide them with extra training. Similarly, if students were admitted because they had shown high levels of general reasoning ability but had lower grades and achievement test scores, then we would want them to know that we thought they had the ability to succeed but that they would need to work harder than other students to do so.[4] This is exactly what happened to many students from small high schools who were admitted to competitive universities because tests like the earlier versions of the *SAT* gave them the opportunity to do so. Indeed, I'd bet there are more than a few of us in this room today.

Conclusions

The allocation of opportunity in society inevitably involves tradeoffs. Decisions that best accomplish one goal may impede the attainment of another, equally cherished outcome. Mental tests have long been recognized as providing one important source of information for college

[4] Note that just as a high level of physical fitness cannot overcome a complete lack of training in a sport, so will high scores on a more fluid reasoning test typically not overcome a lack of knowledge and skill in the domain. This is why I emphasize the importance of students having attainted at least a *moderate* level of knowledge and skill in the domain.

admissions. But tests that best serve the function of measuring prior accomplishment may not be the best measures of future promise. The late Lee Cronbach observed that strange ironies attend the history of mental testing (Cronbach, 1975). Ability tests were once viewed as the liberators of talent for those not privileged by wealth and social status. Then we discovered that they were not as blind to culture or privilege as their advocates had assumed and that they did not measure innate potential in anyone. So they were replaced in many quarters by tests deemed to be fairer because they measure school learning. The irony, though, is that good measures of school learning can show an even larger advantage for the advantaged than measures of reasoning abilities, but only when the reasoning tests are not strongly tied to school learning too. Reasoning tests thus have a place at the admissions table. It is not at the head of the table as some once thought. Rather, such tests provide a way to supplement grades and other measures of past achievement. This is especially important for those who through choice or circumstance have not participated fully in the academic system, or for anyone who is embarking on a course of study that will require new ways of thinking and responding not captured in measures of past achievement. In other words, prior achievement is often an important aptitude for future learning. But it is never the only aptitude, and sometimes not even the most important aptitude.

References

Bingham, W.V. (1937). Aptitudes and aptitude testing. New York: Harper & Brother.

Bridgeman, B., Burton, N., & Cline, F. (2001). Substituting SAT II: Subject tests for SAT I: Reasoning test: Impact on admitted class composition and quality. College Board Research Report No. 2001-3. NY: College Entrance Examination Board.

Cattell, R. B. (1965). The scientific analysis of personality. Baltimore: Penguin Books.

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.

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. (1975). Five decades of public controversy over mental testing. American Psychologist, 30, 1-14.

Donlon, T. F. & Burton, N. W. (1984). The construct and content validity of the SAT. In T.F. Donlon (Ed.), The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.

Eells, K. (1951). Intelligence and cultural differences. Chicago: The University of Chicago Press.

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.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Lemann, N. (1999). The big test: The secret history of the American meritocracy. NY: Farrar, Straus and Giroux.

Lohman, D. F. (1997). Lessons from the history of intelligence testing. International Journal of Educational Research, 27, 1-20.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.) Handbook of intelligence (2nd ed.)(pp. 285-340). Cambridge, MA: Cambridge University Press.

Lubienski, S. T. (2000). A clash of social class cultures? Students experiences in a discussion-intensive seventh-grade mathematics classroom. Elementary School Journal, 100, 377-403.

Martinez, M. E. (2000). Education as the cultivation of intelligence. Mahwah, NJ: Erlbaum.

Snow, E. E. (1996). Aptitude development and education**.** Psychology, Public Policy, & Law, 2, 536-560.

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). Predicting college grades: An analysis of institutional trends over two decades. Princeton, NJ: Educational Testing Service.

Young, J. W. (1990). Adjusting cumulative GPA using Item Response Theory. Journal of Educational Measurement, 27, 175-186.

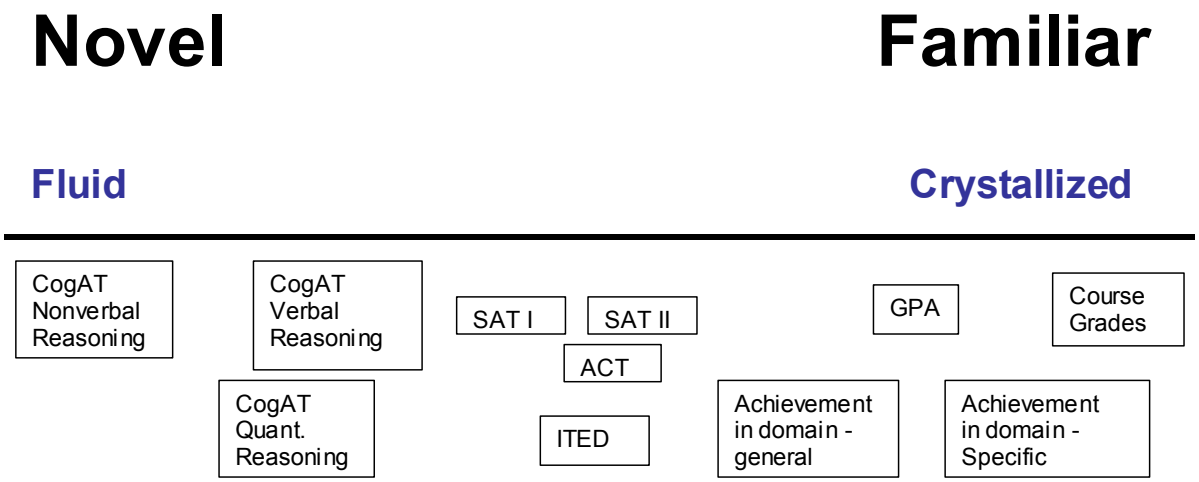

Figure 1. Fluid-Crystallized ability continuum

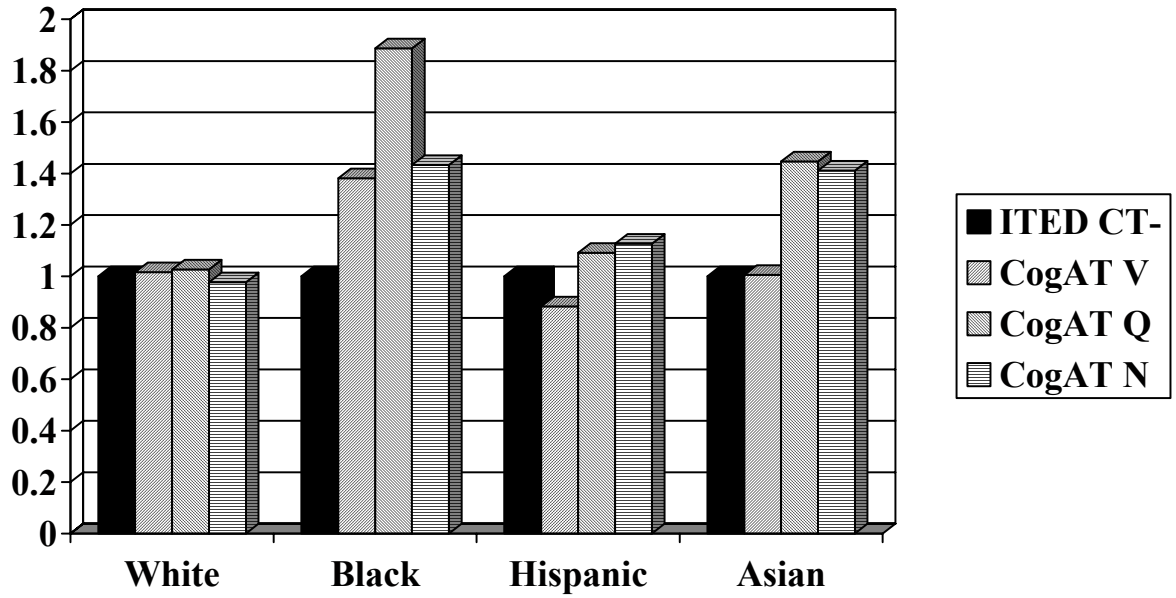

Figure 2. Ratio of the number of students in each ethnic group scoring above the 70th percentile on each test to the number scoring above the 70th percentile on the ITED. By definition, the ratio is fixed at 1.0 for the ITED (black bar). Ratios for the Verbal (V), Quantitative (Q), and Nonverbal (N) batteries of the CogAT are shown in the three hash-marked bars.

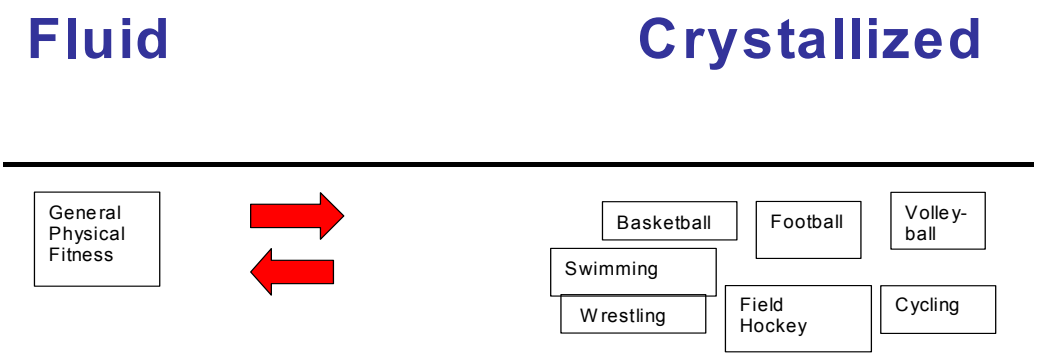

Figure 3. Physical Fitness as aptitude for learning physical skills and as an outcome of participation in such activities.