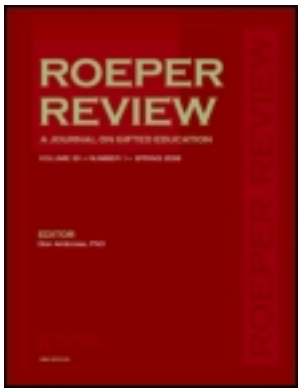

Roeper Review
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uror20

A Big-Picture Thinker Illuminates Identification and Ability: An Interview With David F. Lohman

Suzanna E. Henshon [a]
[a] Florida Gulf Coast University
Published online: 13 Jan 2014.

To cite this article: Suzanna E. Henshon (2014) A Big-Picture Thinker Illuminates Identification and Ability: An Interview With David F. Lohman, Roeper Review, 36:1, 3-10, DOI: 10.1080/02783193.2014.856371

To link to this article: http://dx.doi.org/10.1080/02783193.2014.856371

Roeper Review, 36:3–10, 2014

ISSN: 0278-3193 print / 1940-865X online
DOI: 10.1080/02783193.2014.856371

Routledge
Taylor & Francis Group

AN EVOLVING FIELD

A Big-Picture Thinker Illuminates Identification and Ability: An Interview With David F. Lohman

Suzanna E. Henshon

David F. Lohman is an Emeritus Professor in the Department of Educational Psychology at the University of Iowa, the former Director of Research at the Belin-Blank International Center for Talent Development, and a former staff member of the Iowa Testing Programs. He received his BA in Psychology from Notre Dame in 1972 and his PhD in Educational Psychology from Stanford in 1979. He remained at Stanford as a Lecturer and Research Associate until joining the faculty in the department of Educational Psychology at the University of Iowa in 1981. He served as Chair of the Division of Psychological & Quantitative Foundations in the College of Education from 1993 to 1997. Since 1998, he has authored the *Cognitive Abilities Test*, which is widely used in the United States and other countries to assist in the identification of academically talented children. He is a fellow of the American Psychological Association, the American Psychological Society, and the American Educational Research Association. He is the recipient of numerous awards, including a Fulbright Fellowship, the Iowa Regents Award for Faculty Excellence, and the Distinguished Scholar Award from the National Association of Gifted Children. His research interests include the effectiveness of different curricular adaptations for students who differ in ability or personality, conceptualization and measurement of reasoning abilities, and general issues in the identification and development of talent. E-mail: david-lohman@uiowa.edu

Henshon: *What led you to the field of gifted education?*

Lohman: Like most of the important things that have happened in my life, it was more chance than choice that led me to the field of gifted education.

It all started with a very nasty white rat. The rat was the star performer in a demonstration of the finer principles of behavioral psychology that I was required to complete as undergraduate psychology major. He (and it was a he) had been carefully trained over a period of several weeks. I say "carefully" because the rat was fond of biting any moving object within his field of vision. The day of my presentation finally arrived; I went to retrieve my star pupil from his cage and found him supine, rigor mortis. A few moments of poking with a pencil—which lasted longer than necessary to rouse him—led to the happy conclusion that he was dead but the unhappy conclusion that I would have to start afresh with another nasty white rat. At which point I appealed to my advisor. Surely there was something else that I could do for research project. Better yet, something that did not require establishing a relationship with a rat. And, yes, he had another option. There was a special school near campus where university students helped teach children. I signed up to assist on a project using reinforcement to help control the hyperactive behavior of a rather rambunctious first-grade boy. The boy was challenging but had a great smile. I never went back to rat lab.

The next semester, I enrolled in a course on mental disabilities. The required text by Robinson and Robinson (1965) was one of the few textbooks that I kept long after I had moved on to other topics. (Imagine my surprise many years later when I discovered this was the same Nancy Robinson who has done so much for the field of gifted education.) Eventually I graduated with a degree in psychology. Psychology was not my first or even my main academic interest. I had started out in engineering as a cadet at the U.S. Coast Guard Academy in New London. I went there because a recruiter came to the small Catholic high school that I attended in Great Falls, Montana, and gave a short presentation extolling the virtues of the academy and sailing on the *Eagle*, the 295-foot, three-masted barque used by the Coast Guard for training cadets in the ways of the sea. Had the recruiter not stopped at my school or had I not been asked to attend his talk, I would never have even known about the Coast Guard Academy.[1] Fantasies of sailing aside, the primary appeal of the Academy was that it offered a way to get a college education. I was a good student, but I had no

money for college and no advice on how I might secure it. This, coupled with the prospect of joining a corps of young men with "sound bodies, stout hearts, and alert minds" and sailing on the high seas seemed a much more honorable fate than being drafted into the army for a 2-year tour of Vietnam, which was Option B.

I did well at the Academy. I made great friends, learned to love sailing, but also to abhor the irrational exercise of authority. All was going well until one afternoon during my second year that I tried to convince a friend not to leave. But his arguments were better than mine, and so we both left. He went home to his girlfriend in Kentucky; I went to visit a priest from Stonehill College to discuss the possibility of studying theology. Because I originally hailed from the West, however, the arbitrary rules of the Catholic Church directed that I should pursue my studies at their seminary in the West, which is how I ended up at Notre Dame. Another accident. But I had scored well on the SAT and so I was given a scholarship. The major in psychology was also something of an accident. Thinking that psychology had something to do with Freud, Jung, and like-minded folk that I had read in high school, it seemed like a good foundation for graduate study in theology. Little did I know that Skinner's (1953) *Science and Human Behavior* was the foundational text for the department, which is why the Robinson and Robinson text stood out so strongly. As it turned out, the cognitive psychology I read in the philosophy classes that I took better prepared me for graduate study in the discipline than most of the psychology classes.

By the time I had completed my undergraduate degree, the prospect of lifelong celibacy had seriously dampened my adolescent interest in the priesthood. My interest in theology had also waned. Much of what the nuns had carefully implanted in my brain withstood little serious scrutiny. Indeed, their survey of the world ignored adjacent intellectual continents such as Darwin, the Reformation, and, of course, serious Biblical scholarship. At the time that I was at Notre Dame, at least, the theology department was not being monitored by Rome, and so good teachers encouraged me to step outside of the intellectual bubble in which I had been living. There was also the challenge of making up all of the credits I needed to move from an engineering program to a liberal arts program. I never took anything like a normal course load; every semester I had at least two overloads.[2] But this was 1970 and so the classes were fit to the time. I recall an excellent class in The Philosophy of Law & Revolution and another in Zen Buddhism. The quiet, self-effacing John Dunne's class in the Philosophy of Religion was intellectually transformative. And so by the time I completed my undergraduate degree in psychology, I had received a wonderfully comprehensive undergraduate education that ranged from nuclear physics to metaphysics. However, the BA in psychology was something of a consolation prize.

Henshon: *So how did it happen that you ended up as a psychologist who authors an ability test?*

Lohman: After finishing my undergraduate degree, I returned to Great Falls, Montana, looking forward to some rest and time for reflection. My father tolerated it for a couple of weeks and then gently suggested that I quit loafing and get a job. I believe that there was also some comment about shaving the beard that now adorned my visage. And so I started looking for work. Early in the search I called the father of a high school friend. Floyd McDowell was the Superintendent of the State School for the Deaf and Blind, which also was in Great Falls. On the day I called, he was visiting a new program for multiple-handicapped children that he had created. The program was located 120 miles south in the little town of Boulder, which was the site of the state institution for mentally handicapped children. The only evidence that it was once a thriving mining town were the radon mines visited by people seeking relief from their arthritis. I can only assume that the visitors were not from the Midwest, where radon is free and plentiful in most basements. I also assume that the relationship between radon and lung cancer was as unknown (or ignored) then as the relationship between smoking and lung cancer is now.

Although an undergraduate degree in psychology was not worth much on the job market, my prior experiences with children who had mental and physical disabilities was helpful. And so McDowell offered me a job. I lived in Boulder for 9 months and then escaped north to the main school in Great Falls for a better job. The new job, for which I had little training, required some teaching but mostly administering ability tests to current and prospective students. Because I had a degree in psychology, someone assumed that I knew something about mental testing. I did not. In fact, I remember looking up in the dictionary the word *psychometrics* in the job description (this was before Google). And so, with some assistance from a local school psychologist who had no desire to assume the added burden of testing children at the School for the Deaf & Blind, I learned how to administer most of the major nonverbal tests. This opened my eyes to the field of educational and psychological testing.

Henshon: *Is this what eventually led you to work with Richard Snow at Stanford?*

Lohman: And so when I applied to graduate school, I was keen to work with people whose books I had read—which led me to Lee Cronbach at Stanford. Although I worked with Cronbach occasionally, it was Richard Snow who was my primary mentor. Again, this was not in my playbook. I did not even know who Richard Snow was. He was not only brilliant but also a kind and gentle man. When I arrived on the scene, Cronbach and Snow were finishing their summary and reanalysis of 20 years of research on adapting instruction to individual differences. Cronbach came away from the effort believing that he had entered a hall of mirrors. For example, studies showed that more able learners usually performed best in less structured classes where they were

allowed to discover concepts and principles for themselves. But more anxious students often floundered when structures were removed. So highly anxious, more-able learners needed different instructional support than less anxious, highly able learners. Other personal characteristics and situational variables further complicated the picture—and so the metaphor of a hall of mirrors.

Snow came away with a different view. He felt that the aptitude tests used in this research measured constructs that were poorly understood and thus only distally related to the kinds of thinking and learning that were required of students. And so he had embarked on a multiyear research project funded by the Office of Naval Research that sought better to understand the cognitive processes that distinguished the performance of more- and less-able students on a wide variety of mental tests. The goal was to build cognitive processing models of human abilities and other aptitudes for learning. Bob Sternberg was also a graduate student at Stanford at this time and had embarked on a similar project in his dissertation, which he soon continued to pursue at Yale. Indeed, the Office of Naval Research had funded researchers at many universities who were interested in understanding individual differences in more cognitive psychological terms, not just as collections of traits.

Again, more by accident than anything else, I ended up focusing much of my effort on understanding visual–spatial abilities. At an early meeting of Snow's research team, I was arbitrarily assigned the task of reviewing the correlational research on spatial abilities. Two years later, I had reanalyzed most of the major studies in the field. For many years, the large technical report summarizing those analyses was widely cited—which was nice but somewhat annoying because I had written it during my second year in graduate school! More importantly, though, later debates with others about the value and limitations of nonverbal ability tests were rooted in the extensive research that I had conducted in the measurement and meaning of spatial abilities before, during, and after graduate school.

Snow taught me the importance of considering not only abilities but also motivational, personality, and conative traits and processes when trying to understand individual differences in learning. And, of course, the learning context could make some ways of thinking and responding more helpful than other ways of thinking and responding. This broad view of aptitude came from Snow's early training in industrial–organizational psychology and was reinforced by the studies he and an earlier generation of graduate students had conducted on how best to adapt instruction to the needs and abilities of students. They found that affective traits (such as anxiety, interests, etc.) and conative processes (motivation and volition) mattered as much as general cognitive resources in all complex learning. This more comprehensive view of cognition helped save me from the sort of infatuation with IQ that still characterizes much thinking about giftedness.

I thrived at Stanford. The faculty were excellent; many of the students became future leaders in the field. Scholarship was the coin of the realm and so arguments from authority carried little weight. We were encouraged to think independently but rigorously, to dig deeply into the corpus of research in the field. I would get lost in the stacks of the libraries. One day I stumbled across Terman's copy of Binet's (1909) last book *Les idees modern sur les enfants*, annotated with objections and disagreements. Binet saw mental abilities as developable competencies. In the book, he advocated a kind of mental orthopedics that might assist struggling learners. Terman was not enthused; in fact, he disliked the book so much that he never translated it or asked others to do so. The Stanford professors that I worked with were much less dogmatic, which I appreciated, having seen dogmatism up close for too many years. To this day, the colleagues that I most enjoy—such as David Lubinski at Vanderbilt, Phillip Ackerman at Georgia Tech, Jon Eric Gustafsson at Gothenburg—are people whose scholarship extends well into the past. That many in the field of gifted education have not studied Galton or Binet or E. L. Thorndike or other early leaders in the field surely contributes to the persistence of well-meaning but erroneous beliefs about the nature of intellectual competence.

I was delighted when Snow asked if I would stay on at Stanford after completing the dissertation. We had just received a substantial new grant to pursue our research on individual differences in abilities, motivation, and conation. Then one day in May—long after the faculty recruiting season had ended—a former classmate called me. She never called before; never called again. "There is a job at Iowa that looks like you," she said. Three months later, after another series of accidents and coincidences, I found myself adapting to life in the Midwest.

Henshon: *How did you end up working with Thorndike and Hagen on the Cognitive Abilities Test? Weren't they at Columbia?*

Lohman: In the mid 1990s, it was once again something completely unforeseen that interposed itself in the path of my career, this time moving it directly into the field of gifted education. By that point in my career, I had scaled the academic ladder from assistant to associate to full professor. Shortly after the last promotion, I agreed to serve as Chair of the Division of Psychological & Quantitative Foundations until such time as anyone else wanted the job. Duties of the chair put me in much closer contact with my colleagues in the Iowa Testing Programs (ITP), which is where the Iowa Assessments (formerly the elementary-level *Iowa Test of Basic Skills* and high school level the *Iowa Tests of Educational Development*) are developed. Unbeknownst to me, the publisher of the Iowa tests was looking for someone to assume responsibility for Thorndike and Hagens' *Cognitive Abilities Test*, which is the group-administered

ability test that is conormed with Iowa tests. R. L. Thorndike had passed away during the development of the fifth edition of the test and Betty Hagen was anxious to find someone who could assume responsibility for the sixth edition. My colleagues in ITP suggested my name to the publisher, who asked if I would consider authoring the sixth edition of the test.

This was not an easy decision. Group ability tests do not have an unblemished reputation, especially among those who specialize in administering tests individually. And so before making the decision I did something that I have since discovered that very few people do: I read the manuals. Thorndike and Hagen set the standard for excellence in the development of ability tests. Reading their manuals taught me things about testing that I did not suspect that I did not know. And so I agreed to take on the task.

In many ways, working on CogAT brought together disparate strands of my professional life. Particularly important were my early experiences in testing children, the research that I and my colleagues had conducted on the processes students used on tests similar to subtests in the CogAT batteries, my familiarity with research on the important role of reasoning abilities in guiding instructional adaptations, and, most importantly, the extensive summary of research on both trait and process understandings of intelligence that Snow and I had written for the third edition of *Educational Measurement* (Snow & Lohman, 1989). There was much that I did not know about test development, but I had the great good fortune of advice from Dr. Hagen and, more locally, from my colleague H. D. Hoover in the Iowa Testing Programs.

Once we had completed Form 6 of CogAT, I discovered that I was expected to go out and tell people about it. As a boy, I stuttered terribly, and so public speaking has never been high on my list of favorite activities (see Lohman, 1994). And it was then that I found myself addressing groups of educators who work with gifted children. To my surprise, I had something to say to them—not just because of the test but because of my many years enmeshed in cognitive differential psychology. And when people asked questions that I could not answer, more often than not I would examine the data that we had collected over the years on CogAT and the Iowa tests in an effort to answer. Probably more than anything else, the habit of looking for empirical evidence rather than simply arguments from experts has shaped my understanding of the field.

Henshon: *Can you give some examples of this?*

Lohman: Here is a simple example. How often should we retest children with CogAT? Although I could find sentences that Thorndike or Hagen had written on the issue, I had no firm sense of how much scores on the tests would be expected to change from year to year. And so I examined longitudinal studies using CogAT, the Iowa tests, and other ability and achievement tests and, with the assistance of a graduate student, wrote up my findings and submitted them to one the journals in gifted education (see Lohman & Korb, 2006). The paper explains the causes of the ancient concept of regression towards the mean, which was named and first documented by Galton (1886). Galton actually called it "regression toward mediocrity," which suggests a less positive view of the average person than was advanced by his Belgian and French colleagues. They considered the mean of the distribution as best representing *l'homme moyen* or the ideal type. For them, deviations from this Platonic ideal type were aberrations; for Galton, it was just the opposite, or at least for deviations from the rightmost tail of the distribution (Lohman, 1997).

The causes of regression include but go well beyond errors of measurement. Indeed, anything that lowers the correlation between two sets of status scores increases the amount of regression. And although regression toward the mean is nonexistent at the mean, it increases with every step away from the mean. Thus, the student who receives a very high status score (such as IQ or percentile rank) on one test is unlikely to obtain an equally high score on a second test, but only if there are no practice effects to mask regression.[3] That the field still tolerates policies that permanently assign kindergarten children to special schools (often primarily on the basis of a single test) is inexcusable. As an aside, one of the reviewers of our paper (who was surely a professor at another university) said that he did not believe in regression to the mean, which is probably why the policies are alive and healthy.

Like the regression example, much of my recent work has been didactic—not so much making new discoveries but rather reframing things that we have long known about the measurement of abilities but ignore because they do not comport with our intuitive beliefs. Other examples include the large impact of practice and coaching on ability test scores (E. L. Thorndike, 1919); the difference between culture loading and language loading of tests (Anastasi, 1937); the need for more than figural–spatial item types on nonverbal tests, as on the Army Beta (Lohman & Gambrell, 2012); the imprecision of even high correlations for predictions about individuals (Lohman, 2003; Taylor & Russell, 1939); the difference between the seemingly pure latent variable that a test helps define in a factor analysis and the much messier test score that we use (Spearman, 1904a, 1904b); and the importance of opportunity to learn when making inferences about talent or ability (E. L. Thorndike, Bregman, Cobb, & Woodyard, 1926). Like naïve theories in physics, these beliefs do not change unless specifically addressed. Even then, many find change difficult or impossible (Lohman, 2006).

Henshon: *In some of your papers, you have argued that the field might be better served if we were to abandon the term* gifted*. Why?*

Lohman: In recent years, much of my effort has been devoted to helping people understand how their assumptions about giftedness—and even the word *gifted*—can stand in the way of achieving greater ethnic, cultural, and social

class diversity in their programs. Although I much prefer the broader term *aptitude* to *gifted* or even *intelligence*, I realize that very few people share the comprehensive and carefully nuanced understanding of aptitude that Snow championed. And so I have encouraged people to speak of talent identification and development, and refer them to the work of Gagné (2005) and, more recently, of Subotnik, Olszewski-Kubilius, and Worrell (2011). Unlike the negatively valenced labels in the field of mental disabilities that have a half-life of about 20 years, few educators in the field if gifted education and even fewer parents seem willing to part with the positively valenced term *gifted*. I understand why some—especially those who have struggled to achieve recognition for the needs of academically advanced children or whose very job carries the adjective—would be reluctant to part with the term. However, I do think that school-based programs would not only be better received but face fewer obstacles if they did. Importantly, they would run into fewer roadblocks as they attempt to serve a broader swath of society. Many of the children who were served by the traditional gifted program could be well served by single subject acceleration, especially in hierarchically organized domains such as mathematics and science. Increasingly, Internet-based courses and projects offer possibilities for education that not only build on students' interests but allow collaboration with students in other schools. The talent identification and development specialists who currently teach pull-out classes for the handful of very bright students would then have both license and time to encourage, assist, and direct bright students who do not currently excel to the same degree as those who mainly need acceleration.

Part of my dissatisfaction with the *gifted* label is that it does not challenge people to move beyond unidimensional theories of intelligence that expect gifted children to excel in all domains. This is another example where examining the data effected important changes in my thinking. Once again, it was a question from a user: "Why not just use the overall, three-battery composite score on CogAT when screening for gifted children?" The overall composite is the most reliable score, and it is also most highly correlated with g.

In this case, Thorndike and Hagen were very clear: They admonished users not to screen for gifted children with the overall composite score because it ignores important differences in score profiles. Indeed, only a minority of students (actually about 40%) obtain a relatively flat score profile across the three CogAT batteries. These are the students whose performance is well summarized by the overall composite. For the remaining 60%, however, the composite score misses important information on strengths or weaknesses. This much we knew without looking too carefully at the data. But then we examined the relative frequency of different score profiles by ability level (Lohman, Gambrell, & Lakin, 2008). We discovered that the most able children were at least five times more likely than other children to have an extreme weakness in one area.[4] The opposite held for the least able children: they were much more likely than other children to show an extreme strength in one area. The implication is that although profiles matter at all ability levels, they matter most at the extremes of the distribution. And if we ignore them—for example, by using composite score for identification—then we eliminate many very talented children.

Henshon: *So what are some of the implications for schools that use group tests to screen for talented children?*

Lohman: Screening with any test (such as a typical nonverbal test or even the CogAT 7 Screening Form) that gives only one score can have the same effect. These effects can be mitigated by setting a very liberal cut score on the screening test and then using a placement test battery that gives a dependable profile of students' reasoning abilities and achievements in verbal, quantitative, and spatial domains. Even when the screening and placement tests are highly correlated, the screening test should nominate at least three times as many students as will eventually be admitted or served by the program. If correlations are lower—for example, when a nonverbal screening test is followed by a multiscore individually administered ability test or even a group-administered achievement test—then the cut off score on the screening test should be even more liberal. Else, many highly capable students will not even be considered for admission or placement. Again, this is an example of something that I learned by examining empirical and simulated data (Lohman, 2012b). Even high correlations are much less accurate for making predictions about individual cases than we expect (see Lohman, 2003).

Henshon: *You have also argued that schools should make better use of local norms. Why?*

Lohman: Another unhelpful consequence of the term *gifted* is that it encourages—even demands—reliance on IQs or percentile ranks based on national norms. Although good national norms are useful, they ignore local variations in the distribution of ability. However, the need for special programming at the local level depends on the discrepancy between students' current levels of cognitive or academic development and that of their classmates, not that of all other students in the nation. In many schools, a student with a national percentile rank of 80 can be severely underchallenged. Because schools vary widely in the average ability and achievement of their students, policies that require all students to attain the same level of excellence on a nationally normed test result in schools in which no children are served by the program and other schools in which a substantial fraction of the children are labeled gifted. Local norms eliminate both of these problems (Lohman, 2012c). Although local norms can offer a useful first step, even local norms cannot control for large differences in opportunity to learn among students within the school.

Henshon: *Why is opportunity to learn so important?*

Lohman: At root, all inferences about ability or talent make strong assumptions about similarities in opportunity

to learn. Individual differences in rate or depth of learning can indicate talent. In any domain, children with talent for a particular kind of learning will typically learn in a few trials what otherwise similar children take many trials to learn. Inferences about intellectual ability are thus always judged by the unusualness of a child's performance relative to some larger group of individuals that we assume have had similar opportunities to develop abilities that we observe. This means that the intellectual abilities of students who live in poverty, who have irregular or poor schooling, or who have comparatively less experience with the language of instruction (or testing) can be underestimated when their behavior is compared only to that of other students who have been on the planet for the same number of years and months. Test fairness thus extends beyond judgments about the constructs measured by a test or statistical analyses of performance on them. It must also include fairness of using a particular comparison group when making inferences about the unusualness of the observed performance, even on tests that have been designed to reduce the impact of language and schooling. This is not difficult to do, although it does require abandoning the comforting but illusory notion that a single normative perspective will suffice.[5]

Henshon: *What are some of the things you are working on now?*

Lohman: The final example is one that has emerged from my efforts to help schools maintain the integrity of their assessment programs in a world in which it can no longer be assumed that students are similarly prepared for the ability tests that we give them. Because ability tests measure traits that, although not fixed, are comparatively stable, many erroneously assume that the tests themselves are impervious to external influences. This is not the case. Ability tests are sophisticated but relatively fragile instruments. Nonverbal tests that attempt to reduce the overt impact of language and education are particularly sensitive to practice and coaching. Even without feedback, retest gains of 5 to 10 IQ points on such tests are commonly observed. Practice with feedback and deliberate coaching can produce even larger gains, especially for more able students.

Since Thorndike (1919) first documented the large effects of practice and coaching on ability tests, users have avoided the problem by keeping secret the contents of the tests. Aside from rare instances of cheating, children approached the test with no special preparation. This is no longer the case. The recent proliferation of practice materials sold over the Internet and of coaching schools that operate in many urban areas has seriously undermined the fairness of both group- and individually administered ability tests when test scores are used for high-stakes admissions decisions. For a price, savvy parents with resources can virtually assure their child a high score and thus of placement in the gifted program.

Test developers, test users, and educational policy makers can make changes to current policies and practices that address this problem. Test developers can provide practice activities for all students, thereby partially releveling the playing field; they can develop multiple forms of their tests, thereby reducing the advantage gained by access to the items on any one form; and on some tests they can report indices that caution users when analyses of item performance suggest coaching or cheating.[6] Although helpful, none of these changes will completely solve the problem. The most important change can only be made by those who set the policies that specify how scores on ability tests are used in the talent identification process.

An IQ score of 130 (or national percentile rank of 97) on an ability test has long been required for admission to G&T programs in many states. Although experts have cautioned against this sort of high-stakes use of ability test scores, the convenience of the policy has outweighed its negative consequences. In some states, one of the more desired outcomes has been an increase in funding: every child whose scores exceed the standard brings additional resources to the program. A less desired outcome is that students with artificially high scores are often less prepared for the demands of the G&T program than classmates who did not practice, scored just below the cut, and were not admitted. Even more problematic is the decrease in the diversity of the student population that is served by the program. The parents of poor and underserved minority students typically are not the purchasers of either practice tests or admission to test preparation classes. Unequal practice thus not only invalidates scores for those who receive it but effectively unravels the often considerable efforts that programs have made to diversify the population of students that they serve.

Reducing the stakes attached to scores on the ability test is the key to fairer and more defensible policies. The ability test score needs to be one of the more lenient criteria in the selection procedure rather than the most restrictive criterion. A good example is Renzulli's (2005) recommendation to use ability tests to help identify a large talent pool. Children whose needs can be addressed are assigned to different types and levels of intervention using evidence on interests, accomplishments, and scholastic achievement. The scores on the ability test are set as a lower bound, typically using local stanines or local percentile ranks that make eligible for consideration the top 15%–25% of the student population. This seemingly liberal standard actually comports well with what we know about the plurality of abilities and their imperfect relationships with achievements in different domains.

Setting a more generous cut score on the ability test simultaneously reduces the stakes placed on the ability test and increases the importance of evidence of achievement and creative production. Reducing the stakes on the ability test also reduces both the need for and value of external practice on the test, thereby preserving the validity of the scores students obtain on it. Emphasizing accomplishments and creative contributions in science, mathematics, and the arts encourages substantive preparation in these domains rather than unproductive cramming for an ability test.

Henshon: *Conclusions? Thoughts for the future of the discipline?*

Lohman: Strange ironies attend the history of mental testing (Cronbach, 1975). National norms are most useful when students hail from different parts of the nation. However, all of the large, multistate talent development programs search for talented children using what are essentially local norms on the SAT, ACT, or other ability tests. On the other hand, most schools search for gifted children using IQ scores or national percentile ranks. There is a larger irony in the way in which we interpret scores on ability and achievement tests. The interpretive frameworks for ability and achievement tests have switched: The measurement of achievement now emphasizes national comparisons while the measurement of ability increasingly emphasizes local comparisons. For both ability and achievement, though, multiple perspectives are needed, with different normative comparisons best aligned with different kinds of inferences.

When properly interpreted, ability tests are most useful for the have-nots rather than the haves. If a child in second grade is doing fifth-grade mathematics, then she needs more challenging instruction in mathematics. At best, a measure of her quantitative (or general) reasoning abilities provides useful ancillary information. Rather, the ability test is most useful for the child who has talent but has not had much opportunity to develop those talents. This is why ability tests tend to be much more useful for young children than adolescents and for the have-nots than the haves. It is thus no small irony that many educators concerned with bias place ability tests at the top of their list. But even without further consideration of opportunity to learn, mean differences between children who differ in ethnicity or social class are much smaller on well-constructed ability tests than on well-constructed achievement tests.[7] Eliminating ability tests and moving to a world in which challenging academic opportunities are provided only to those who currently excel academically is, as Cronbach and Snow (1977) put it, only slightly less conservative than a social system that allocates opportunity on the basis of hereditary social status.

I would like to predict the demise of Terman's IQ-based definition of giftedness. But I know that this view appeals to most of the lay public and even many educators. Even those who endorse a multidimensional view of giftedness often resort to a single IQ-like score because it is convenient and, in some cases, necessary given limitations on the number of students who can be served by the program.

Change is difficult in most fields, especially large, complex fields such as education. Often, the threat of negative consequences is a more effective motivator than impassioned arguments of reformers. Perhaps the twin challenges posed by the need to serve a broader swath of society than is currently served by many programs and the inability to rely so heavily on ability test scores will motivate changes in the ways the field uses and abuses ability test scores. And at the end of the day, only those changes that are reflected in federal, state, and local educational policies matter.[8] My colleagues at the Institute for Research and Policy on Acceleration (IRPA) have shown how this can be done. Although many educators were convinced by the arguments for allowing students to accelerate, few school districts had policies that allowed it. IRPA's sample policy has now been widely used in the United States and abroad. We need a similarly well-articulated policy template for the use of ability tests that not only conform to our best understanding of the multidimensionality of academic giftedness but also confront the corrupting influence that the high stakes uses of scores on ability tests has had on the integrity of the test scores in the age of the Internet.

Ability tests can provide critical information on academic talent. This information is most useful for those who, through age, experience, circumstance, or choice, have not developed high levels of competence in some academic or cognitive domain. My personal experience and the experiences of many of my academic colleagues confirm this claim. We grew up in families that were short on economic resources but long on opportunities for work. Ability and achievement tests gave us access to the educational resources we needed. But ability tests provide only one piece of the picture. Effective learning requires different ensembles of cognitive, conative, and affective aptitudes for learning in different content domains, at different stages of learning, and by different methods of instruction. For example, the volition control strategies that I developed as a boy to persist in the pursuit of goals that seemed beyond my reach were later essential for capitalizing on the educational opportunities that good test scores provided. Now that I have run the race, though, will these habits be as useful? Can I develop a more measured pace to life in the slower lane? I hope so. But regardless of what the future holds, it has been an amazing journey. I am painfully aware, though, that many others who came from similar circumstances did not have the repeated good fortune that I had. And as the cost of a good education rises, the disparity between childhood promise and adult attainment will only increase. Now more than ever, education needs those who are committed to identifying and then assisting children who have the requisite aptitudes for high-level accomplishment but not the economic and educational resources needed to develop them.

NOTES

1. The only other recruiter that year was from Columbia, which crossed my path later when I agreed to assume responsibility for Thorndike and Hagens' *Cognitive Abilities Test*. Both were on the faculty at Teachers College.
2. I had only one elective in the $4\frac{1}{2}$ years it took to complete the degree. I opted for Greek.
3. Oddly, most practitioners ignore the substantial literature on the effects of practice and coaching on ability tests. E. L. Thorndike was so worried about these effects that he abandoned tests that tried to present somewhat novel problems and relied instead on well-practiced educational tasks (see R. M. Thorndike & Lohman, 1990).

4. When compared specifically to students who have a median stanine score of 5, the increase in the number with an extreme weakness is actually eightfold.
5. The WISC-IV Spanish is a good example. In addition to the traditional index scores that compare the student to all other bilingual Spanish-speaking children in the nation, a percentile rank can be reported that controls for amount of education in U.S. schools and a second percentile rank that also controls for parental education. The proper normative perspective to use depends on the kind of inference that one wants to make from the test scores. See also the chapter on Fairness in the *CogAT Form 7 Research and Development Guide* (Lohman, 2012a).
6. This is difficult to do unless the test has many items and the proportion of items that appear to indicate cheating is high. Results that are sufficiently accurate for group work are typically not sufficiently accurate for defending the inference that an individual has cheated or been coached on parts of the test.
7. Although achievement gaps on NAEP have declined in recent years, full-population estimates still show Black–White differences of approximately 1 SD. Differences on ability tests are now $\frac{1}{3}$ to $\frac{1}{2}$ SD and even smaller if family income is controlled.
8. See, for example, Brown, Avery, VanTassel-Baska, Worley, and Stambaugh (2006).

REFERENCES

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

Binet, A. (1909). *Les idées modern sur les enfants [Modern ideas about children]*. Paris, France: E. Flammarion.

Brown, E., Avery, L., VanTassel-Baska, J., Worley, B., & Stambaugh, T. (2006). A five-state analysis of gifted education policy. *Roeper Review, 29*, 11–23.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116–127.

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

Gagné, F. (2005). From gifts to talents. The DMGT as a developmental model. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 93–112). Cambridge, England: Cambridge University Press.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Lohman, D. F. (1994). Spatially gifted, verbally, inconvenienced. In N. Colangelo, S. G. Assouline, & D. L Ambroson (Eds.), *Talent development: Vol. 2. Proceedings from the 1993 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 251–264). Dayton, OH: Ohio Psychology Press.

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

Lohman, D. F. (2003). Tables of prediction efficiencies. Retrieved from http://faculty.education.uiowa.edu/david-lohman/home

Lohman, D. F. (2006). Beliefs about differences between ability and accomplishment: From folk theories to cognitive science. *Roeper Review*, *29*, 32–40.

Lohman, D. F. (2012a). *Cognitive Abilities Test, Form 7: Research and development guide*. Rolling Meadows, IL: Riverside.

Lohman, D. F. (2012b). Decision strategies. In S. L. Hunsaker (Ed.), *Identification: The theory and practice of identifying students for gifted and talented education services* (pp. 217–248). Mansfield Center, CT: Creative Learning Press.

Lohman, D. F. (2012c). Nontraditional uses of traditional measures. In C. M. Callahan & H. Hertberg-Davis (Eds.), *Fundamentals of gifted education* (pp. 274–284). New York, NY: Taylor & Francis/Routledge.

Lohman, D. F., & Gambrell, J. (2012). Use of nonverbal measures in gifted identification. *Journal of Psychoeducational Assessment*, *30*, 25–44.

Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science Quarterly*, *50*, 269–282.

Lohman, D. F., & Korb, K. (2006). Gifted today but not tomorrow? Longitudinal changes in ITBS and CogAT scores during elementary school. *Journal for the Education of the Gifted*, *29*, 451–484.

Renzulli, J. S. (2005). *Equity, excellence, and economy in a system for identifying students in gifted education: A guidebook* (RM05208). Storrs, CT: University of Connecticut, The National Research Center on the Gifted and Talented.

Robinson, H. B., & Robinson, N. M. (1965). *The mentally retarded child*. New York, NY: McGraw-Hill.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York, NY: Macmillan.

Spearman, C. (1904a). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–291.

Spearman, C. (1904b). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*, 3–54.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical validity of tests in selection: Discussion and tables. *Journal of Applied Psychology*, *23*, 565–578.

Thorndike, E. L. (1919). Tests of intelligence; Reliability, significance, susceptibility to special training, and adaptation to the general nature of the task. *School and Society*, *9*, 189–195.

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

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

AUTHOR BIO

Dr. Suzanna E. Henshon graduated from The College of William & Mary and teaches creative writing at Florida Gulf Coast University. She is the author of seven published books, including *Spiders on the Ceiling* (2006) and *King Arthur's Academy: Descriptive and Narrative Exercises* (2007). E-mail: shenshon@fgcu.edu