

Educational Testing: Measuring and Remediating Achievement Gaps

JAEKYUNG LEE

Abstract

Achievement gaps, as measured by standardized tests, are inextricably related to educational goals, standards, norms, and benchmarks for student learning outcomes. I revisit conventional approaches to educational testing to measure achievement gaps—norm-referenced, criterion-referenced, and potential-referenced tests. I explore and discuss a paradigm shift from “passive” tests to “responsive” tests that promotes the diagnosis and remediation of achievement gaps. Particularly, I propose an environment-referenced approach to testing with the specification of desired learning opportunities and environment conditions that enable students to meet upgraded achievement norms, standards, and benchmarks.

BACKGROUND ISSUES

Standardized tests play increasingly important roles in monitoring and shaping American education. Since the Coleman Report in the 1960s brought attention to racial and socioeconomic gaps based on standardized achievement test results (Coleman *et al.*, 1966), the National Assessment of Educational Progress (NAEP) served as the nation’s report card to help monitor national progress in educational equity (Jencks & Phillips, 1998). While most states relied more on basic skills test in the 1970s, *A Nation at Risk* report of 1983 called for an end to the minimum competency testing movement (Amrein & Berliner, 2002). While the focus of education policy shifted from equity to excellence, the target achievement goal of high-stakes testing policy also shifted from minimum competency to proficiency. The *No Child Left Behind Act* of 2001 (NCLB) was aimed at accomplishing high academic standards for all students and closing their achievement gaps among racial and social groups. Schools and teachers are held accountable for their students’ test results as monitored by the state and local education agencies. Educational policy changes have been increasingly influenced

by international math and science achievement test results that reveal the gap between American students and their peers in other industrial nations (Baker, 2003; Husén & Tuijnman, 1994; Lee, 2001).

Achievement gaps as measured by standardized tests are inextricably related to educational goals, standards, norms, and benchmarks for student learning outcomes. There are three conventional approaches to educational testing to measure achievement gaps—norm-referenced, criterion-referenced, and potential-referenced tests. Performance-driven, external accountability policy mandates, with increasingly higher standards and expectations for both student and teacher performance, demand understanding and tackling the sources of achievement gaps. While recent changes in accountability and testing policies have provided educators with access to an abundance of student achievement data, there are no clear framework and evidence-based guidelines about how to use data for guiding instruction and improving student learning (Hamilton *et al.*, 2009). While the trend of increased online and digital education environment facilitates personalized learning with real-time data collection and feedback with ongoing measurement of learning inputs, processes and outcomes, analyzing such big data for integrated educational decision making poses new challenges (US Department of Education, 2012).

In light of these issues and challenges, I revisit conventional testing approaches that narrowly focus on measuring outcomes without full consideration of examinees' characteristics and learning environment. First, I explain multiple approaches to educational test design for measuring and reporting achievement gaps. In the context of education policy changes and technology upgrades, I further explore and discuss a paradigm shift toward integrated measurement of student learning environment and achievement gaps, particularly, an environment-referenced approach to testing with the specification of desired learning opportunities and engagement conditions that enable students to meet upgraded achievement norms, standards, and benchmarks. This paradigm shift signifies transition from "passive" tests that simply measure and report achievement gaps to "responsive" tests that assess learning environment needs and guide educational interventions for closing the achievement gaps.

DEFINITION AND MEASUREMENT OF ACHIEVEMENT GAPS

Achievement gaps arise from the discrepancy between desired goals and current status. Performance goals can be set with reference to either criteria or norms (or a combination of both). Achievement gaps change over the course of child development and education. This requires dynamic, rather static, views about the measurement and analysis of achievement gaps.

Traditionally, tests have been designed to measure achievement gaps for norm-referenced versus criterion-referenced evaluation purposes. However, this distinction depends on interpretation of the test results more than on the test itself. The two types can be viewed as a continuum rather than a clear-cut dichotomy. Test publishers often try to make the test versatile, with possible dual interpretations (Linn & Gronlund, 2000). Both norms and criteria are not fixed, and they are likely to change over time. As Table 1 shows, there were upward shifts in the level and rigor of performance goals, both norms and standards over the past several decades at the national level.

Previous studies have often investigated the achievement gap without questioning the validity of traditional reference groups or standards for comparison. Students representing particular racial (e.g., White students) or social groups (e.g., parents with college education) are often chosen as the reference groups based on stereotypes and these students' historically superior performance. Race and ethnicity variables, however, are crude proxies for the educational needs and academic risks that vary significantly among individual students within such broadly defined groups. While a criterion-referenced approach (e.g., state's performance standards for math proficiency) has become more popular in the past two decades of standards-based education reform, this approach often results in unrealistically high standards with underfunded mandates. Furthermore, a strict criterion-referenced approach does not consider the varied needs of students and also does not outline multiple pathways for diverse groups of students to reach common standards. Further, both norms and criteria tend to be uniformly applied to the entire age or grade cohort group of students.

Why do some groups of students have greater academic success than others? According to John Carroll's model of school learning (Carroll, 1963), the amount of learning is a function of students' aptitude and prior knowledge that determines the time needed for learning, as well as learning opportunity that determines the time available for learning and perseverance (engagement) that affects time spent on learning. In this sense, the achievement gap occurs when actual achievement is significantly below expected achievement based on one's best ability and effort under optimal environmental conditions. The notion of "plasticity" in human development processes and outcomes is well established (Baltes, Lindenberger, & Staudinger, 2006; Oyserman, Bybee, & Terry, 2006; Vygotsky, 1978). Instead of common benchmarks, academic performance goals can be customized with reference to potential for each individual student (i.e., capacity for the future achievement).

Table 1
 Testing Approaches for the Assessment of Achievement Gaps against Upgraded Norms, Standards, and Predictions

Trend	Norm-referenced Testing (NRT)	Criterion-referenced Testing (CRT)	Potential-referenced Testing (PRT)
Old	Gap relative to national average achievement or White average achievement	Gap against minimum competency standard for high school graduation	Gap against achievement prediction based on standardized test measures of aptitude (IQ, SAT) or prior achievement (GPA, ACT)
New	Gap relative to international average achievement or high-performing Asian average achievement	Gap against proficiency standard for college readiness (e.g., common core standards)	Gap against achievement prediction based on family background characteristics as well as aptitude and prior achievement
Emerging –augmented by “environment-referenced” testing (ERT) approach	Gap relative to international average achievement of model nations with desired learning opportunities and engagement conditions	Gap against proficiency standard for college readiness with desired learning opportunities and engagement conditions	Gap against achievement prediction based on both observable and unobservable potential under desired learning opportunities and engagement conditions

In the following sections, each of these different testing approaches for measurement of academic achievement gap is discussed. Although my differentiation of testing approaches in this article relies on the basis of comparison and reference, tests can be also classified in different ways such as formative vs. summative assessments based on testing purposes (Lee & Lee, 2013). Here I focus on the case of standardized tests of which traditional purpose was often limited to performance evaluation, certification, and/or selection, but I argue that its roles can and should be expanded to more diagnostic and remedial functions as augmented by environmental scan of learning conditions that underlie the achievement results.

NORM-REFERENCED TESTING

Norm-referenced tests (NRTs) compare an examinee's performance to that of other examinees; this practice is often called "grading on the curve." Standardized examinations such as the SAT are NRTs. The goal is to rank the set of examinees so that decisions about their opportunity for success (e.g., college entrance) can be made. Norms refer to performances by defined groups on particular tests at a particular time. Norms are used to give information about one's performance relative to a standardization sample (e.g., mean, median). Norms are obtained by getting the distribution of test scores from a normative sample, whether tests measure achievement or aptitude. Achievement tests measure what a student has learned, and aptitude tests measure the ability (readiness) to learn new tasks. For this reason, NRTs are designed

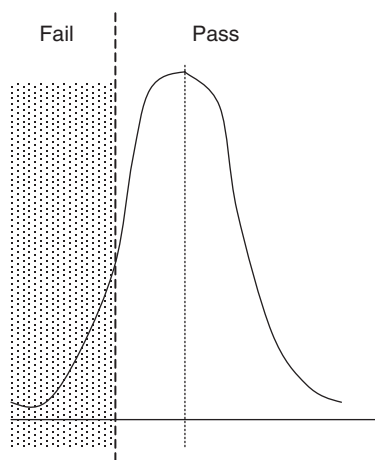


Figure 1 Hypothetical distributions of academic achievement with performance norms and standards cut score for performance standard.

to maximize discrimination among examinees and produce the wide spread of achievement distribution in a bell curve format (Figure 1).

For example, a conventional norm-referenced view for college admissions relies on relative performance of students; for example, highly selective colleges may select out the top 5–10% of students based on college entrance exam scores or high school ranks. While the selection of cutoff is arbitrary, this high academic talent group often has special opportunities and privileges such as gifted/talented education, advanced placement courses, and merit-based college scholarship. For example, the top 10% college admissions policy in Texas guarantees automatic admissions into a state flagship university for high school students whose high school grade ranks fall in the top 10% of the graduating class. This cutoff clearly represents the norm-referenced threshold; in this case, the norm used for sorting is not statewide norms but local school norms.

There have been changes in reference groups in the analysis of racial achievement gaps. Comparisons between Black and White students are different from comparisons between Black and Asian-American students, and give different insights into unique issues of domestic inequalities and discriminations for different minority groups in the United States. However, these conventional within-country group comparisons do not address new global economic challenges that put even White students (i.e., the traditional majority group in the United States) or Asian-American students (i.e., the highest performing group in the United States) at a relative disadvantage in comparison with native Asians on an international scale. This kind of international achievement gap (as opposed to the domestic achievement gap) will more directly determine students' odds of college admissions in the globalized college marketplace and job employment in the globalized labor market.

CRITERION-REFERENCED TESTING

Criterion-referenced tests (CRTs) compare each examinee's performance to a predefined set of criteria or a standard. The goal with these tests is to determine whether the candidate has the demonstrated mastery of a certain skill or set of skills. These results are usually "pass" or "fail" and are used in making decisions about job entry, certification, or licensure (see cut score for performance standard in Figure 1). A national board medical examination is an example of a CRT; either the examinee has the skills to practice the profession, in which case he or she is licensed, or does not. Multiple standard-setting methods have been devised and used to determine pass/fail test cutscores (see Cizek, 1993; Jaeger, 1989). For the criterion-referenced approach to measuring the achievement gap in K-12

education, both national and international assessments provide well-defined student performance standards in the era of standards-based education reform. The Trends in International Math and Science Study (TIMSS) *High* achievement benchmark requires student competency of knowledge/skills application for problem-solving across content domains (Martin, Mullis, & Foy, 2008). The NAEP *Proficient* achievement level demands that students are able to demonstrate competency over challenging subject matter, including conceptual understanding, application of knowledge to real-world situations, and analytical skills (NAGB, 2001).

More recently, the federal *Race to the Top* policy pushed for adopting common core standards for college and career readiness and performance-based teacher evaluations based on student achievement of the standards. Conventional views of college readiness have focused narrowly on students' first-year college placement or performance as predicted by college admissions tests for college-bound students only, which remain to be separated from recent K-12 standards-based educational assessment movement toward improved college readiness for all students (ACT, 2010; Lee, 2012a). There is also the expectation that a typical average college-bound student (as opposed to a high-performing or gifted student) in American high schools will achieve a grade of B or higher. For example, as part of the K-16 state policy initiatives, the Georgia HOPE scholarship program guarantees a state scholarship for students with a B high school average of GPA, where only college preparation courses count as part of the B average requirement.

An example of criterion-referenced formative assessments is curriculum-based measurement (CBM) that employs repeated, frequent measurements of student performance in basic skills by classroom teachers (Stecker, Fuchs, & Fuchs, 2005). CBM makes comparison of student progress with reference to long-term (year-long) curriculum goals and is often applied in Response to Intervention (RTI) context: CBM can help identify students in need of interventions, decide which level of intervention is most appropriate, and determine if an intervention is successful (Mellard & Johnson, 2008). This tiered, continuous improvement approach facilitates differentiated instruction and adaptive intervention.

POTENTIAL-REFERENCED TESTING

Potential-referenced tests (PRTs) compare each examinee's performance to a predicted level of achievement or potential. While the norms or criteria are uniformly identified for all students through observed achievement distributions (e.g., mean scores) or agreed benchmarks (e.g., cut scores), the PRT approach is specific to individual students. However, potential is not directly observable and more difficult to operationalize and measure. For this

reason, potential may be measured as a range rather than a point on a continuous scale of achievement. The achievement goal may be set low within this boundary of potential or high enough to push the boundary out further. The capacity to which students come to realize their full potential determines the degree of deficits relative to goals. Under this framework, the achievement gap is treated as unrealized potential (underachievement).

This individualistic, PRT approach was used in the conventional measurement of learning disabilities through a combination of intelligence testing and academic achievement testing. The resulting information on discrepancy between IQ and achievement is used to determine whether a child's academic performance is commensurate with his or her cognitive ability (Figure 2). If a child's cognitive ability is much higher than his or her academic performance, the student is often diagnosed with a learning disability. Although the discrepancy model has dominated school practices in the past, its validity and utility have been questioned (Aaron, 1995). Another popular application of PRT approach in education is computer-adaptive testing (CAT). It enables the design and administration of customized tests that can match the difficulty level of test items to the (latent) ability level of examinees based on the item response theory (IRT). CAT successively selects questions for the purpose of maximizing the precision of the examination based on what is known about the examinee from previous questions (Weiss & Kingsbury, 1984). Although CAT serves to improve efficiency and flexibility in standardized testing itself, it does not address any environmental causes

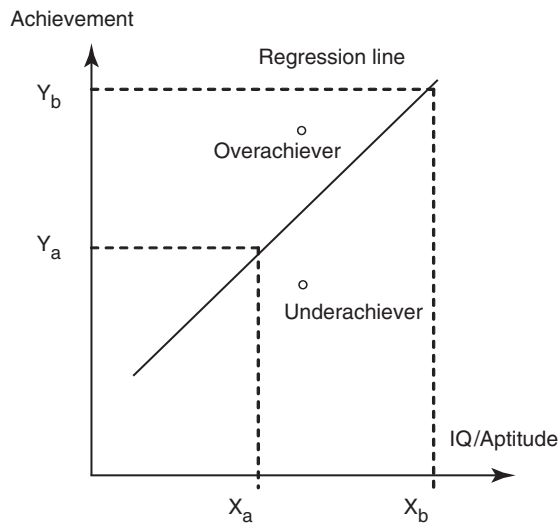


Figure 2 Hypothetical relationship between IQ/aptitude and achievement for identification of overachiever and underachiever.

of observed variations in academic ability and achievement. The issue is not simply how to adapt tests to varying student ability/achievement at the time of testing, but rather how to change educational environment for the improvement of student ability/achievement itself.

PARADIGM SHIFT: ENVIRONMENT-REFERENCED TESTING FOR GAP DIAGNOSIS AND REMEDIATION

Each of these three approaches has its own distinctive features and utilities. However, the distinction becomes blurred if test design encompasses all three aspects of achievement gap information. Further, an emerging and potentially fruitful area of interdisciplinary research is linking achievement expectations (norms, criteria, and potential) to educational environment, including both home and school learning conditions. Here, I would call it the “environment-referenced” testing (ERT) approach. It can augment any of the aforementioned three approaches to provide information for achievement gap diagnosis and remediation. This cross-referencing means developing achievement norms, references, standards, and expectations consistent with desirable learning conditions (e.g., access to quality curriculum and teachers, parental support, student engagement) and incorporating those conditions into construction of corresponding norms, standards, and predictions.

Although researchers disagree on the best measure of school/teacher quality and the significance of school/teacher effects, a great deal of evidence suggests systematically positive effects of instructional resources and teacher quality on academic achievement (Ferguson, 1991; Hedges, Laine, & Greenwald, 1994). Previous studies also demonstrated both disparities and inadequacies in terms of school funding and teacher quality for disadvantaged minority groups (Darling-Hammond, 2000; Hanushek, Kain, & Rivkin, 2001; Ingersoll, 1996; Lankford, Loeb, & Wyckoff, 2002). Poor minority students are often double-bound by problems with less adequate instructional resources and less qualified teachers in their schools along with challenges posed by their relatively disadvantaged home learning environment. Further, students’ academic engagement is crucial for translating learning opportunities into achievement (Finn, 1989; NRC, 2004). Therefore, the new framework of educational testing must address two components: (i) engagement to learning (ETL) and (ii) opportunity to learn (OTL). As summarized in Table 1, the operational definition of achievement gap is extended to embed the concept and measurement of learning environment gap.

This ERT approach requires upfront specification of desired learning conditions and thus addresses the loophole of black box approaches as seen in the value-added models (VAMs) of teacher and school performance evaluation.

The VAM often predicts student achievement based on a combination of prior achievement, personal and family background characteristics that influence the current assessment results beyond the control of teachers and/or schools. Positive values indicate performance that is better than expected based on students' background information, whereas negative values indicate underperformance. Those residual values, the gap between actual and predicted achievement scores, are used to evaluate unique contributions of teachers and/or schools to student learning (McCaffrey, Lockwood, & Koretz, 2004). However, the VAM approach leaves desired schooling conditions unspecified and thus makes the results vulnerable to distortion of instructional processes such as teaching to the test.

RECOMMENDATIONS FOR FURTHER READING

In the context of NRT, applying the environment-referenced approach means specifying desired learning conditions for selected normative groups. For example, achievement gap may be defined as a gap relative to international average achievement of model nations with specification of desired learning conditions, including learning opportunities and engagement efforts. An example of similar approach in health can be found in the World Health Organization (WHO) "Model of International Reference for Child Growth" (Garza, 2006; Garza & de Onis, 2004). The new model adds "prescriptive" and "international" aspects to child growth norms, strengthening of advocacy for child health. First of all, their new model attempts to prescribe how children should grow rather than describe how children grow. It involves using sample selection criteria consistent with health promotion recommendations (e.g., breastfeeding norms, standard pediatric care, and nonsmoking requirements). This new global standard emphasizes the notion that all humans are equal and that environmental differences rather than genetic endowments are the principal determinants of disparities in physical growth (Garza & de Onis, 2004).

In the context of CRT, environment-referenced gaps can be measured against proficiency standards such as standards for college readiness, with specification of desired learning conditions and opportunities. The focus of school finance reform has shifted from equity to adequacy in the midst of the performance-based educational accountability movement (Bartman, 2002; Clune, 1994; Ladd, Chalk, & Hansen, 1999). There are several approaches to measuring school funding adequacy, including an empirical observation of successful districts/schools and econometric cost function analysis (Ladd *et al.*, 1999). However, states that adopted high-stakes testing policies during the past decade often failed to improve key school resources and address funding inequalities (Lee & Wong, 2004). Given the moving target

of academic performance standards, more research is needed to tackle the issue of how adequate and equitable are the distributions of school and teacher resources (e.g., per pupil expenditures and qualified teachers) to help different groups of students meet common and rigorous proficiency standards (Lee, 2012b).

In the context of PRT, environment-referenced, gaps can be measured against predictions based on observable and unobservable potential with specification of desired learning conditions. With desirable learning environment conditions in place, one can predict potential achievement level, and assess how much an individual student or a group of students realizes the potential. This potential-referenced approach also has implications for the measurement of racial achievement gaps. Ferguson (2007) discusses three different conceptions of racial bias in teacher expectations of student learning potential: unconditional race neutrality, weak conditional race neutrality, and strong conditional race neutrality. "Strong" conditional race neutrality starts with the assumption that Black and White children are born with the same potential, and that there is no distinction at birth, yet that disparities in potential may develop as children grow older under different environmental conditions. What is recommended for policy and practice here is not simply making the assumption but rather collecting data on the gaps in learning environment. In the context of college admissions, this environment-referenced approach may ask colleges to collect information on the history of applicants' educational opportunities and incorporate that information into interpretations of test results and admissions decision; this approach is different from adopting test-optional admissions policy or using alternative measures (Soares, 2011). It also calls for developing and validating test items that are not biased against or sensitive to students from disadvantaged learning environment backgrounds (AERA, APA, & NCME, 1999).

REFERENCES

- Aaron, P. G. (1995). Differential diagnosis of reading disabilities. *School Psychology Review, 24*(3), 345–360.
- ACT (2010). *College readiness standards for EXPLORE, PLAN and the ACT*. Retrieved from www.act.org.
- American Educational Research Association, American Psychological Association, & National Council on Educational Measurement (1999). *Standards for educational and psychological testing*. Retrieved from <http://www.apa.org/science/standards.html>.
- Amrein, A. L., & Berliner, D. C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved June 14, 2003 from <http://epaa.asu.edu/epaa/v10n18/>

- Baker, D. P. (2003). Should we be more like them? Reflections on causes of cross-national high school achievement differences and implications for American educational reform policy. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 309–325). Washington, DC: Brookings Institution.
- Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2006). Life span theory in developmental psychology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 569–664). New York, NY: Wiley.
- Bartman, K. D. (2002). Public education in the 21st century: How do we ensure that no child is left behind? *Temple Political & Civil Rights Law Review*, 12(1), 95–119.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723–733.
- Cizek, G. J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Clune, W. H. (1994). The shift from equity to adequacy in school finance. *Educational Policy*, 8(4), 376–394.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, A. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Retrieved from <http://epaa.asu.edu/epaa/v8n1/>
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117–142.
- Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation*, 28, 465–498.
- Ferguson, R. F. (2007). *Toward excellence with equity: An emerging vision for closing the achievement gap*. Cambridge, MA: Harvard Education Press.
- Garza, C. (2006). New growth standards for the 21st century: A prescriptive approach. *Nutrition Reviews*, 64(5), S55–S59.
- Garza, C., & De Onis, M. (2004). Rationale for developing a new international growth reference. *Food and Nutrition Bulletin*, 25(Supplement 1), S5–14.
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009–4067)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2001). *Why public schools lose teachers* (Working Paper 8599). Cambridge, MA: National Bureau of Economic Research.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5–14.
- Husén, T., & Tuijnman, A. (1994). Monitoring standards in education: Why and how it came about. In A. Tuijnman & T. N. Postlethwaite (Eds.), *Monitoring the standards of education* (pp. 1–21). Oxford, UK: Pergamon.

- Ingersoll, R. (1996). The problem of under-qualified teachers in American secondary schools. *Educational Researcher*, 28(2), 26–37.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (pp. 485–514). New York, NY: Macmillan.
- Jencks, C., & Phillips, M. (Eds.) (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Ladd, H. F., Chalk, R., & Hansen, J. S. (Eds.) (1999). *Equity and adequacy in education finance: Issues and perspectives*. Washington, DC: National Academy Press.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62.
- Lee, J. (2001). School reform initiatives as balancing acts: Policy variation and educational convergence among Japan, Korea, England and the United States. *Education Policy Analysis Archives*, 9(13). Retrieved from <http://epaa.asu.edu/epaa/v9n13.html>
- Lee, J. (2012a). College for all: Gaps between desirable and actual P-12 math achievement trajectories for college readiness. *Educational Researcher*, 41(2), 43–55.
- Lee, J. (2012b). Educational equity and adequacy for disadvantaged minority students: School and teacher resource gaps toward national math proficiency standard. *Journal of Educational Research*, 105(1), 64–75.
- Lee, J., & Lee, Y. S. (2013). Effects of testing. In J. Hattie & E. Anderman (Eds.), *International guide to student achievement* (pp. 416–418). New York, NY: Routledge.
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797–832.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- McCaffrey, D. F., Lockwood, J. R., & Koretz, D. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67–101.
- Mellard, D. F., & Johnson, E. (2008). *RTI: A practitioner's guide to implementing response to intervention*. Thousand Oaks, CA: Corwin Press.
- National Assessment Governing Board (2001). *National assessment of educational progress achievement levels 1992–1998 for mathematics*. Washington, DC: Author.
- National Research Council and the Institute of Medicine (2004). *Engaging schools: Fostering high school students' motivation to learn*. Committee on Increasing High School Students' Engagement and Motivation to Learn. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, 91, 188–204.

- Soares, J. A. (Ed.) (2011). *SAT wars: The case for test-optional college admissions*. New York, NY: Teachers College Press.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819.
- US Department of Education, Office of Educational Technology (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: Author.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361–375.

JAEKYUNG LEE SHORT BIOGRAPHY

Jaekyung Lee is the Dean and Professor of Education at the University at Buffalo, SUNY. He has a PhD in education from the University of Chicago. He is a fellow of American Educational Research Association (AERA) and a former fellow of the Center for Advanced Study in the Behavioral Sciences at Stanford University. He is the author of the book, *The Testing Gap: Scientific Trials of Test-Driven School Accountability Systems for Excellence and Equity*. His research focuses on the issues of achievement gaps and educational equity.

RELATED ESSAYS

- Economics of Early Education (*Economics*), W. Steven Barnett
- Shadow Education (*Sociology*), Soo-yong Byun and David P. Baker
- The Organization of Schools and Classrooms (*Sociology*), David Diehl and Daniel A. McFarland
- Expertise (*Sociology*), Gil Eyal
- Evolutionary Approaches to Understanding Children’s Academic Achievement (*Psychology*), David C. Geary and Daniel B. Berch
- The Evidence-Based Practice Movement (*Sociology*), Edward W. Gondolf
- Retrieval-Based Learning: Research at the Interface between Cognitive Science and Education (*Psychology*), Ludmila D. Nunes and Jeffrey D. Karpicke
- The Impact of Learning Technologies on Higher Education (*Psychology*), Christopher S. Pentoney *et al.*
- Curriculum as a Site of Political and Cultural Conflict (*Sociology*), Fabio Rojas
- Education in an Open Informational World (*Educ*), Marlene Scardamalia and Carl Bereiter