

education review
a journal of book reviews

[reseñas educativas \(Spanish\)](#)
[resenhas educativas \(Portuguese\)](#)

This review has been accessed **483** times since May 3, 2008

Lee, Jaekyung. (2007). *The Testing Gap: Scientific Trials of Test-Driven School Accountability Systems for Excellence and Equity*. Charlotte, North Carolina: Information Age Publishing.

Pp. xi + 190 \$40 ISBN 978-1-59311-748-1

**Reviewed by Monique Herbert, Saad Chahine, & Ruth A. Childs
Ontario Institute for Studies in Education, University of Toronto**

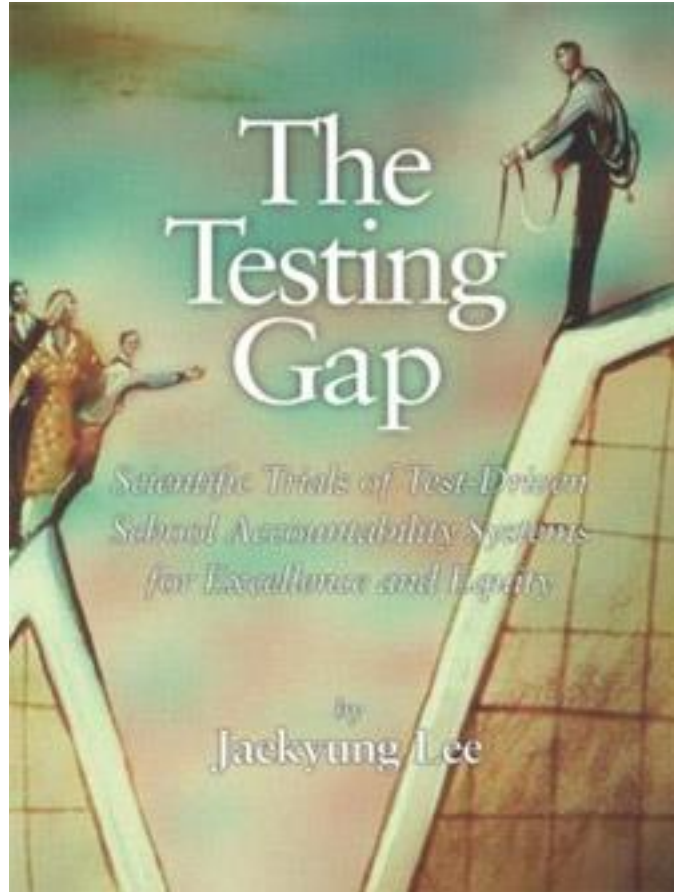
May 3, 2008

In *The Testing Gap: Scientific Trials of Test-Driven School Accountability Systems for Excellence and Equity*, Jaekyung Lee summarizes a decade of his own research on trends in United States student achievement, as revealed by state assessments and the National Assessment of Educational Progress (NAEP). Lee's research, some of it conducted for The Civil Rights Project at Harvard University and much of it funded by the US Department of Education, compared, when the data allowed, trends for individual states and for Black, Hispanic, and White students and poor and non-poor students.

The passage of the No Child Left Behind Act of 2001 (NCLB) during that decade of research gave Lee an additional comparison: before and after NCLB. Of particular interest to Lee were the questions, "Did NCLB change the trends in overall achievement?" and "Did NCLB change the gaps between groups?" In addition, he investigated whether changes might differ between states that had weak accountability systems before NCLB and those that had strong accountability systems.

In Part I of *The Testing Gap*, Lee provides a theoretical and critical overview of accountability systems in the US, the complexities of creating policies for specific accountability goals and of implementing those policies, and the perceived tensions between the goals of excellence and equity. Lee uses this overview to draw attention to what he believes to be one of the major issues related to accountability policy, research, and practice: the dearth of scientifically-based research to "better inform and evaluate educational policy" (p. 21).

This is a crucial point and highlights the uniqueness of the contribution Lee makes to the vast literature on testing through his research. In fact, the US is one of the few countries that is likely to yield the kinds of scientific trials Lee describes. Many countries, because of the structure of their education systems (e.g., small numbers of jurisdictions, lack of funding, no country-wide educational policies), provide limited opportunities for large-scale research on the effects of educational policies. However, the structure of education in the US – the 50 states independently control their educational systems, but the federal government provides some funding and can put conditions on that funding – makes the US a unique laboratory for educational experiments.





Jaekyung Lee

Lee must be commended for dedicating an entire chapter in Part I to the technical threats that may exist in state and national assessments and in accountability systems. This chapter focuses on the use of test scores as an indicator of student achievement and “the reliability and validity of AYP [Annual Yearly Progress] indicators as measures of school accountability” (p. 68). The importance of considering the psychometric properties (i.e., validity, reliability, fairness) of test scores and indicators, such as AYP, derived from those scores cannot be emphasized too strongly. Using the states of Maine and Kentucky as examples of weak and strong accountability systems, Lee not only presents several ways to assess the validity and reliability of these accountability systems and the assessments they use, but also discusses the pitfalls associated with the statistical analyses. For example, even though Kentucky and Maine both use assessment frameworks modeled closely on the NAEP framework and the correlation between those states’ assessment results and NAEP results are relatively high, this information does not provide us with sufficient evidence of criterion validity. The use, here and elsewhere, of tables and figures to present the analysis results

should help readers who are not familiar with the complex statistical analyses Lee uses to understand the results of those analyses.

In Part II, Lee presents a comprehensive analysis of the average achievement trends and achievement gap trends pre- and post-NCLB. The results of hierarchical linear modeling (HLM) analyses reveal that, at both the national and state levels, the current trends are unlikely to lead to 100% of students demonstrating proficiency by 2014, the goal set by NCLB. The most striking and persistent findings are the absence of gains in reading achievement and the continuing presence of racial and socioeconomic inequalities for both reading and mathematics. Lee cautions, however, that these findings are based on only four years of post-NCLB data and that NAEP standards are higher than state standards.

In the final section, Part III, Lee discusses what he sees as the implications of his research for educational policy and the design and implementation of accountability systems. Lee picks up again the exploration of “accountability for excellence” and “accountability for equity” started in Part I. The contrast between “learning gaps” and “testing gaps” is also discussed, although it is unfortunate that this important contrast is introduced only in the concluding chapter, when it could have been used effectively throughout the book.

As Lee emphasizes, *The Testing Gap* summarizes the results of much-needed research on trends in student achievement, results that have the potential to inform evaluations of the current accountability systems and the design of future systems. Without diminishing Lee’s contributions, however, it is important to note some limitations of both the research described and the description itself. For example, the impact of NCLB and state accountability on the achievement gap is difficult to interpret without more information about the programming and curricular changes that occurred within states as a result of or at about the same time as NCLB. In addition, much of the research is limited to a few states. Of particular note is the focus on Maine in some of the analyses, which is understandable, as Lee was a professor at the University of Maine for several years. However, the sparseness of Maine’s population means that small schools in Maine are often very small and far from libraries and other community supports. Whether the factors Lee finds to be related to educational achievement and improvement in Maine schools will generalize to other states is questionable.

The book appears to have been hastily assembled, with much of it drawn from other reports (e.g., Lee, 2004, 2006). Unfortunately, its many grammatical and typographical errors and unnecessarily complex sentences make it a difficult read. The lack of clear definitions of terms that are important

to Lee's arguments also add to the difficulty. One notable example is the apparently interchangeable use of the terms test-driven accountability and performance-based accountability. These and other terms are likely to have different meanings in different contexts. The inclusion of definitions would especially increase the book's usefulness for international researchers.

Despite these quibbles, Lee presents a worthwhile and welcomed contribution to the literature on large-scale assessment and accountability systems. The content of this book warrants attention by educators, researchers, test developers, practitioners, and policy makers.

References

Lee, J. (2004, April 7). How feasible is Adequate Yearly Progress (AYP)? Simulations of school AYP "Uniform Averaging" and "Safe Harbor" under the No Child Left Behind Act. *Education Policy Analysis Archives*, 12(14). Retrieved December 16, 2007 from <http://epaa.asu.edu/epaa/v12n14>

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: The Civil Rights Project at Harvard University. Retrieved December 16, 2007 from http://www.civilrightsproject.ucla.edu/research/esea/nclb_naep_lee.pdf

About the Reviewers

Monique Herbert is a Ph.D. candidate in Developmental Psychology and Education. Her research interests include large-scale testing with particular emphasis on the administration and use of large-scale assessments and examining patterns of missing data. Her current research examines the developmental trajectories exhibited by young people who engage in delinquent behaviors, especially the intraindividual changes (transitions) that occur.

Saad Chahine is a Ph.D. candidate in Developmental Psychology and Education. His research focuses on large-scale assessment and classroom assessment in relation to accountability policy. He is currently involved in a large project exploring the formative use of large-scale assessment data.



Monique Herbert

Ruth Childs is an associate professor in Human Development and Applied Psychology. Her research focuses on practical psychometric issues that arise in large-scale assessments, such as the advantages and disadvantages of matrix sampling, alternative approaches to scoring tests, and the impact of missing data treatments. She is currently investigating how teachers decide whether to comply with test preparation and test administration guidelines.

Copyright is retained by the first or sole author, who grants right of first publication to the *Education Review*.



[Editors: Gene V Glass, Kate Corby, Gustavo Fischman](#)

~ [ER home](#) | [Reseñas Educativas](#) | [Resenhas Educativas](#) ~

~ [overview](#) | [reviews](#) | [editors](#) | [submit](#) | [guidelines](#) | [announcements](#) | [search](#) ~



Ruth Childs