



Fordham Institute's Pretend Research

by Richard P. Phelps

The Thomas B. Fordham Institute has released a report, *Evaluating the Content and Quality of Next Generation Assessments*,¹ ostensibly an evaluative comparison of four testing programs, the Common Core-derived SBAC and PARCC, ACT's Aspire, and the Commonwealth of Massachusetts' MCAS.² Of course, anyone familiar with Fordham's past work knew beforehand which tests would win.

This latest Fordham Institute Common Core apologia is not so much research as a caricature of it.

1. Instead of referencing a wide range of relevant research, Fordham references only friends from inside their echo chamber and others paid by the Common Core's wealthy benefactors. But, they imply that they have covered a relevant and adequately wide range of sources.
2. Instead of evaluating tests according to the industry standard *Standards for Educational and Psychological Testing*, or any of dozens of other freely-available and well-vetted test evaluation standards, guidelines, or protocols used around the world by testing experts, they employ "a brand new methodology" specifically developed for Common Core, for the owners of the Common Core, and paid for by Common Core's funders.
3. Instead of suggesting as fact only that which has been rigorously evaluated and accepted as fact by skeptics, the authors continue the practice of Common Core salespeople of attributing benefits to their tests for which no evidence exists
4. Instead of addressing any of the many sincere, profound critiques of their work, as confident and responsible researchers would do, the Fordham authors tell their critics to go away—"If you don't care for the standards...you should probably ignore this study" (p. 4).

Dr. Richard P. Phelps is editor or author of four books: *Correcting Fallacies about Educational and Psychological Testing* (APA, 2008/2009); *Standardized Testing Primer* (Peter Lang, 2007); *Defending Standardized Testing* (Psychology Press, 2005); and *Kill the Messenger* (Transaction, 2003, 2005), and founder of the Nonpartisan Education Review (<http://nonpartisaneducation.org>).

5. Instead of writing in neutral language as real researchers do, the authors adopt the practice of coloring their language as so many Common Core salespeople do, attaching nice-sounding adjectives and adverbs to what serves their interest, and bad-sounding words to what does not.

1. Common Core's primary private financier, the Bill & Melinda Gates Foundation, pays the Fordham Institute handsomely to promote the Core and its associated testing programs.³ A cursory search through the Gates Foundation web site reveals \$3,562,116 granted to Fordham since 2009 expressly for Common Core promotion or "general operating support."⁴ Gates awarded an additional \$653,534 between 2006 and 2009 for forming advocacy networks, which have since been used to push Common Core. All of the remaining Gates-to-Fordham grants listed supported work promoting charter schools in Ohio (\$2,596,812), reputedly the nation's worst.⁵

The other research entities involved in the latest Fordham study either directly or indirectly derive sustenance at the Gates Foundation dinner table:

- the Human Resources Research Organization (HumRRO),⁶
- the Council of Chief State School Officers (CCSSO), co-holder of the Common Core copyright and author of the test evaluation "Criteria."⁷
- the Stanford Center for Opportunity Policy in Education (SCOPE), headed by Linda Darling-Hammond, the chief organizer of one of the federally-subsidized Common Core-aligned testing programs, the Smarter-Balanced Assessment Consortium (SBAC),⁸ and
- Student Achievement Partners, the organization that claims to have inspired the Common Core standards⁹

The Common Core's grantees have always only hired their own well-subsidized grantees for evaluations of

their products. The Buros Center for Testing at the University of Nebraska has conducted test reviews for decades, publishing many of them in its annual *Mental Measurements Yearbook* for the entire world to see, and critique. Indeed, Buros exists to conduct test reviews, and retains hundreds of the world's brightest and most independent psychometricians on its reviewer roster. Why did Common Core's funders not hire genuine professionals from Buros to evaluate PARCC and SBAC? The non-psychometricians at the Fordham Institute would seem a vastly inferior substitute, ...that is, had the purpose genuinely been an objective evaluation.

2. A second reason Fordham's intentions are suspect rests with their choice of evaluation criteria. The "bible" of North American testing experts is the *Standards for Educational and Psychological Testing*, jointly produced by the American Psychological Association, National Council on Measurement in Education, and the American Educational Research Association. Fordham did not use it.¹⁰

Had Fordham compared the tests using the *Standards for Educational and Psychological Testing* (or any of a number of other widely-respected test evaluation standards, guidelines, or protocols¹¹) SBAC and PARCC would have flunked. They have yet to accumulate some the most basic empirical evidence of reliability, validity, or fairness, and past experience with similar types of assessments suggest they will fail on all three counts.¹²

Instead, Fordham chose to reference an alternate set of evaluation criteria concocted by the organization that co-owns the Common Core standards and co-sponsored their development (Council of Chief State School Officers, or CCSSO), drawing on the work of Linda Darling-Hammond's SCOPE, the Center for Research on Educational Standards and Student Testing (CRESST), and a handful of others.^{13,14} Thus, Fordham compares SBAC and PARCC to other tests according to specifications that were designed for SBAC and PARCC.¹⁵

The authors write “The quality and credibility of an evaluation of this type rests largely on the expertise and judgment of the individuals serving on the review panels” (p.12). A scan of the names of everyone in decision-making roles, however, reveals that Fordham relied on those they have hired before and whose decisions they could safely predict. Regardless, given the evaluation criteria employed, the outcome was foreordained regardless whom they hired to review, not unlike a rigged election in a dictatorship where voters’ decisions are restricted to already-chosen candidates.

PARCC and SBAC might have flunked even if Fordham had compared tests using all 24+ of CCSSO’s “Criteria.” But Fordham chose to compare on only 14 of the criteria.¹⁶ And those just happened to be criteria mostly favoring PARCC and SBAC.

Without exception the Fordham study avoided all the evaluation criteria in the categories:

“Meet overall assessment goals and ensure technical quality”,

“Yield valuable reports on student progress and performance”,

“Adhere to best practices in test administration”,
and

“State specific criteria”¹⁷

What types of test characteristics can be found in these neglected categories? Test security, providing timely data to inform instruction, validity, reliability, score comparability across years, transparency of test design, requiring involvement of each state’s K-12 educators and institutions of higher education, and more. Other characteristics often claimed for PARCC and SBAC, without evidence, cannot even be found in the CCSSO criteria (e.g., internationally benchmarked, backward mapping from higher education standards, fairness).

The report does not evaluate the “quality” of tests, as its title suggests; at best it is an alignment study. And, naturally, one would expect the Common Core

consortium tests to be more aligned to the Common Core than other tests. The only evaluative criteria used from of the CCSSO’s Criteria are in the two categories “Align to Standards—English Language Arts” and “Align to Standards—Mathematics” and, even then, only for grades 5 and 8.

Nonetheless, the authors claim, “The methodology used in this study is highly comprehensive” (p. 74).

The authors of the Pioneer Institute’s report *How PARCC’s False Rigor Stunts the Academic Growth of All Students*,¹⁸ recommended strongly against the official adoption of PARCC after an analysis of its test items in reading and writing. They also did not recommend continuing with the current MCAS, which is also based on Common Core’s mediocre standards, chiefly because the quality of the grade 10 MCAS tests in math and ELA has deteriorated in the past seven or so years for reasons that are not yet clear. Rather, they recommend that Massachusetts return to its effective pre-Common Core standards and tests and assign the development and monitoring of the state’s mandated tests to a more responsible agency.

Perhaps the primary conceit of Common Core proponents is that the familiar multiple-choice/short answer/essay standardized tests ignore some, and arguably the better, parts of learning (the deeper, higher, more rigorous, whatever).¹⁹ Ironically, it is they—opponents of traditional testing content and formats—who propose that standardized tests *measure everything*. By contrast, most traditional standardized test advocates do not suggest that standardized tests can or should measure any and all aspects of learning.

Consider this standard from the Linda Darling-Hammond, et al. source document for the CCSSO criteria:

“Research: Conduct sustained research projects to answer a question (including a self-generated question) or solve a problem, narrow or broaden the inquiry when appropriate, and demonstrate

understanding of the subject under investigation. Gather relevant information from multiple authoritative print and digital sources, use advanced searches effectively, and assess the strengths and limitations of each source in terms of the specific task, purpose, and audience.”²⁰

Who would oppose this as a learning objective? But, does it make sense as a standardized test component? How does one objectively and fairly measure “sustained research” in the one- or two-minute span of a standardized test question? In PARCC tests, this is done by offering students snippets of documentary source material and grading them as having analyzed the problem well if they cite two of those already-made-available sources.

But, that is not how research works. It is hardly the type of deliberation that comes to most people’s mind when they think about “sustained research”. Advocates for traditional standardized testing would argue that standardized tests should be used for what standardized tests do well; “sustained research” should be measured more authentically.

The authors of the aforementioned Pioneer Institute report recommend, as their 7th policy recommendation for Massachusetts:

“Establish a junior/senior-year interdisciplinary research paper requirement as part of the state’s graduation requirements—to be assessed at the local level following state guidelines—to prepare all students for authentic college writing.”²¹

PARCC, SBAC, and the Fordham Institute propose that they can validly, reliably, and fairly measure the outcome of what is normally a weeks- or months-long project in a minute or two.²² It is attempting to measure that which cannot be well measured on standardized tests that makes PARCC and SBAC tests “deeper” than others. In practice, the alleged deeper parts are the most convoluted and superficial.

Appendix A of the source document for the CCSSO criteria provides three international examples

of “high-quality assessments” in Singapore, Australia, and England.²³ None are standardized test components. Rather, all are projects developed over extended periods of time—weeks or months—as part of regular course requirements.

Common Core proponents scoured the globe to locate “international benchmark” examples of the type of convoluted (i.e., “higher”, “deeper”) test questions included in PARCC and SBAC tests. They found none.

3. The authors continue the Common Core sales tendency of attributing benefits to their tests for which no evidence exists. For example, the Fordham report claims that SBAC and PARCC will:

“make traditional ‘test prep’ ineffective” (p. 8)

“allow students of all abilities, including both at-risk and high-achieving youngsters, to demonstrate what they know and can do” (p. 8)

produce “test scores that more accurately predict students’ readiness for entry-level coursework or training” (p. 11)

“reliably measure the essential skills and knowledge needed ... to achieve college and career readiness by the end of high school” (p. 11)

“...accurately measure student progress toward college and career readiness; and provide valid data to inform teaching and learning.” (p. 31)

eliminate the problem of “students ... forced to waste time and money on remedial coursework.” (p. 73)

help “educators [who] need and deserve good tests that honor their hard work and give useful feedback, which enables them to improve their craft and boost their students’ success.” (p. 73)

The Fordham Institute has not a shred of evidence to support any of these grandiose claims. They share more in common with carnival fortune telling than empirical research. Granted, most of the statements refer to future outcomes, which cannot be known with

certainty. But, that just affirms how irresponsible it is to make such claims absent any evidence.

Furthermore, in most cases, past experience would suggest just the opposite of what Fordham asserts. Test prep is more, not less, likely to be effective with SBAC and PARCC tests because the test item formats are complex (or, convoluted), introducing more “construct irrelevant variance”—that is, students will get lower scores for not managing to figure out formats or computer operations issues, even if they know the subject matter of the test. Disadvantaged and at-risk students tend to be the most disadvantaged by complex formatting and new technology.

As for Common Core, SBAC, and PARCC eliminating the “problem of” college remedial courses, such will be done by simply cancelling remedial courses, whether or not they might be needed, and lowering college entry-course standards to the level of current remedial courses.

4. When not dismissing or denigrating SBAC and PARCC critiques, the Fordham report evades them, even suggesting that critics should not read it: “If you don’t care for the standards...you should probably ignore this study” (p. 4).

Yet, cynically, in the very first paragraph the authors invoke the name of Sandy Stotsky, one of their most prominent adversaries, and a scholar of curriculum and instruction so widely respected she could easily have gotten wealthy had she chosen to succumb to the financial temptation of the Common Core’s profligacy as so many others have. Stotsky authored the Fordham Institute’s “very first study” in 1997, apparently. Presumably, the authors of this report drop her name to suggest that they are broad-minded. (It might also suggest that they are now willing to publish anything for a price.)

Tellingly, one will find Stotsky’s name nowhere after the first paragraph. None of her (or anyone else’s) many devastating critiques of the Common Core tests is either mentioned or referenced. Genuine

research does not hide or dismiss its critiques; it addresses them.

Ironically, the authors write, “A discussion of [test] qualities, and the types of trade-offs involved in obtaining them, are precisely the kinds of conversations that merit honest debate.” Indeed.

5. Instead of writing in neutral language as real researchers do, the authors adopt the habit of coloring their language as Common Core salespeople do. They attach nice-sounding adjectives and adverbs to what they like, and bad-sounding words to what they don’t.

For PARCC and SBAC one reads:

“strong content, quality, and rigor”

“stronger tests, which encourage better, broader, richer instruction”

“tests that focus on the essential skills and give clear signals”

“major improvements over the previous generation of state tests”

“complex skills they are assessing.”

“high-quality assessment”

“high-quality assessments”

“high-quality tests”

“high-quality test items”

“high quality and provide meaningful information”

“carefully-crafted tests”

“these tests are tougher”

“more rigorous tests that challenge students more than they have been challenged in the past”

For other tests one reads:

“low-quality assessments poorly aligned with the standards”

“will undermine the content messages of the standards”

“a best-in-class state assessment, the 2014 MCAS, does not measure many of the important competencies that are part of today’s college and career readiness standards”

“have generally focused on low-level skills”

“have given students and parents false signals about the readiness of their children for postsecondary education and the workforce”

Appraising its own work, Fordham writes:

“groundbreaking evaluation”

“meticulously assembled panels”

“highly qualified yet impartial reviewers”

Considering those who have adopted SBAC or PARCC, Fordham writes:

“thankfully, states have taken courageous steps”

“states’ adoption of college and career readiness standards has been a bold step in the right direction.”

“adopting and sticking with high-quality assessments requires courage.”

A few other points bear mentioning. The Fordham Institute was granted access to operational SBAC and PARCC test items. Over the course of a few months in 2015, the Pioneer Institute, a strong critic of Common Core, PARCC, and SBAC, appealed for similar access to PARCC items. The convoluted run-around responses from PARCC officials excelled at bureaucratic stonewalling. Despite numerous requests, Pioneer never received access.

The Fordham report claims that PARCC and SBAC are governed by “member states”, whereas ACT Aspire is owned by a private organization. Actually, the Common Core Standards are owned by two private, unelected organizations, the Council of Chief State School Officers and the National Governors’ Association, and only each state’s chief school officer sits on PARCC and SBAC panels. But, individual states actually have for more say-so if they adopt ACT

Aspire (or their own test) than if they adopt PARCC or SBAC. A state adopts ACT Aspire under the terms of a negotiated, time-limited contract. By contrast, a state or, rather its chief state school officer, has but one vote among many around the tables at PARCC and SBAC. With ACT Aspire, a state controls the terms of the relationship. With SBAC and PARCC, it does not.²⁴

Just so you know, on page 71, Fordham recommends that states eliminate any tests that are not aligned to the Common Core Standards, in the interest of efficiency, supposedly.

In closing, it is only fair to mention the good news in the Fordham report. It promises on page 8, “We at Fordham don’t plan to stay in the test-evaluation business.”

Endnotes

1. Nancy Doorey & Morgan Polikoff. (2016, February). *Evaluating the content and quality of next generation assessments*. With a Foreword by Amber M. Northern & Michael J. Petrilli. Washington, DC: Thomas B. Fordham Institute.
<http://edexcellence.net/publications/evaluating-the-content-and-quality-of-next-generation-assessments>
2. PARCC is the Partnership for Assessment of Readiness for College and Careers; SBAC is the Smarter-Balanced Assessment Consortium; MCAS is the Massachusetts Comprehensive Assessment System; ACT Aspire is not an acronym (though, originally ACT stood for American College Test).
3. The reason for inventing a Fordham Institute when a Fordham Foundation already existed may have had something to do with taxes, but it also allows Chester Finn, Jr. and Michael Petrilli to each pay themselves two six figure salaries instead of just one.
4. <http://www.gatesfoundation.org/search#q/k=Fordham>
5. See, for example, <http://www.ohio.com/news/local/charter-schools-mispend-millions-of-ohio-tax-dollars-as-efforts-to-police-them-are-privatized-1.596318>; http://www.cleveland.com/metro/index.ssf/2015/03/ohios_charter_schools_ridicule.html; <http://www.dispatch.com/content/stories/local/2014/12/18/kasich-to-revamp-ohio-laws-on-charter-schools.html>; <https://www.washingtonpost.com/news/answer-sheet/wp/2015/06/12/troubled-ohio-charter-schools-have-become-a-joke-literally/>
6. HumRRO has produced many favorable reports for Common Core-related entities, including alignment studies in Kentucky, New York State, California, and Connecticut.
7. CCSSO has received 23 grants from the Bill & Melinda Gates Foundation from “2009 and earlier” to 2016 collectively exceeding \$100 million.
<http://www.gatesfoundation.org/How-We-Work/Quick-Links/Grants-Database#q/k=CCSSO>
8. <http://www.gatesfoundation.org/How-We-Work/Quick-Links/Grants-Database#q/k=%22Stanford%20Center%20for%20Opportunity%20Policy%20in%20Education%22>
9. Student Achievement Partners has received four grants from the Bill & Melinda Gates Foundation from 2012 to 2015 exceeding \$13 million. <http://www.gatesfoundation.org/How-We-Work/Quick-Links/Grants-Database#q/k=%22Student%20Achievement%20Partners%22>
10. The authors write that the standards they use are “based on” the real *Standards*. But, that is like saying that Cheez Whiz is based on cheese. Some real cheese might be mixed in there, but it’s not the product’s most distinguishing ingredient.
11. (e.g., the International Test Commission’s (ITC) *Guidelines for Test Use*; the ITC *Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*; the ITC *Guidelines on the Security of Tests, Examinations, and Other Assessments*; the ITC’s *International Guidelines on Computer-Based and Internet-Delivered Testing*; the European Federation of Psychologists’ Association (EFPA) Test Review Model; the Standards of the Joint Committee on Testing Practices)
12. Despite all the adjectives and adverbs implying newness to PARCC and SBAC as “Next Generation Assessment”, it has all been tried before and failed miserably. Indeed, many of the same persons involved in past fiascos are pushing the current one. The allegedly “higher-order”, more “authentic”, performance-based tests administered in Maryland (MSPAP), California (CLAS), and Kentucky (KIRIS) in the 1990s failed because of unreliable scores; volatile test score trends; secrecy of items and forms; an absence of individual scores in some

cases; individuals being judged on group work in some cases; large expenditures of time; inconsistent (and some improper) test preparation procedures from school to school; inconsistent grading on open-ended response test items; long delays between administration and release of scores; little feedback for students; and no substantial evidence after several years that education had improved. As one should expect, instruction had changed as test proponents desired, but without empirical gains or perceived improvement in student achievement. Parents, politicians, and measurement professionals alike overwhelmingly rejected these dysfunctional tests.

See, for example, **For California**: Michael W. Kirst & Christopher Mazzeo, (1997, December). The Rise, Fall, and Rise of State Assessment in California: 1993-96, *Phi Delta Kappan*, 78(4) Committee on Education and the Workforce, U.S. House of Representatives, One Hundred Fifth Congress, Second Session, (1998, January 21). National Testing: Hearing, Granada Hills, CA. Serial No. 105-74; Representative Steven Baldwin, (1997, October). Comparing assessments and tests. *Education Reporter*, 141. See also Klein, David. (2003). "A Brief History Of American K-12 Mathematics Education In the 20th Century", In James M. Royer, (Ed.), *Mathematical Cognition*, (pp. 175–226). Charlotte, NC: Information Age Publishing. **For Kentucky**: ACT. (1993). "A study of core course-taking patterns. ACT-tested graduates of 1991-1993 and an investigation of the relationship between Kentucky's performance-based assessment results and ACT-tested Kentucky graduates of 1992". Iowa City, IA: Author; Richard Innes. (2003). Education research from a parent's point of view. Louisville, KY: Author. <http://www.eddatafrominnes.com/index.html> ; KERA Update. (1999, January). Misinformed, misled, flawed: The legacy of KIRIS, Kentucky's first experiment. **For Maryland**: P. H. Hamp, & C. B. Summers. (2002, Fall). "Education." In P. H. Hamp & C. B. Summers (Eds.), A guide to the issues 2002–2003. Maryland Public Policy Institute, Rockville, MD. <http://www.mdpolicy.org/docLib/20051030Education.pdf>; Montgomery County Public Schools. (2002, Feb. 11). "Joint Teachers/Principals Letter Questions MSPAP", Public Announcement, Rockville, MD. <http://www.montgomeryschoolsmd.org/press/index.aspx?pagetype=showrelease&id=644>; HumRRO. (1998). Linking teacher practice with statewide assessment of education. Alexandria, VA: Author. <http://www.humrro.org/corpsite/page/linking-teacher-practice-statewide-assessment-education>

13. [http://www.ccsso.org/Documents/2014/CCSSO Criteria for High Quality Assessments 03242014.pdf](http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf)
14. A rationale is offered for why they had to develop a brand new set of test evaluation criteria (p. 13). Fordham claims that new criteria were needed, which weighted some criteria more than others. But, weights could easily be applied to any criteria, including the tried-and-true, preexisting ones.
15. For an extended critique of the CCSSO Criteria employed in the Fordham report, see "Appendix A. Critique of Criteria for Evaluating Common Core-Aligned Assessments" in Mark McQuillan, Richard P. Phelps, & Sandra Stotsky. (2015, October). *How PARCC's false rigor stunts the academic growth of all students*. Boston: Pioneer Institute, pp. 62-68. <http://pioneerinstitute.org/news/testing-the-tests-why-mcas-is-better-than-parcc/>
16. Doorey & Polikoff, p. 14.
17. MCAS bests PARCC and SBAC according to several criteria specific to the Commonwealth, such as the requirements under the current Massachusetts Education Reform Act (MERA) as a grade 10 high school exit exam, that tests students in several subject fields (and not just ELA and math), and provides specific and timely instructional feedback.
18. McQuillan, M., Phelps, R.P., & Stotsky, S. (2015, October). *How PARCC's false rigor stunts the academic growth of all students*. Boston: Pioneer Institute. <http://pioneerinstitute.org/news/testing-the-tests-why-mcas-is-better-than-parcc/>

19. It is perhaps the most enlightening paradox that, among Common Core proponents' profuse expulsion of superlative adjectives and adverbs advertising their "innovative", "next generation" research results, the words "deeper" and "higher" mean the same thing.
20. The document asserts, "The Common Core State Standards identify a number of areas of knowledge and skills that are clearly so critical for college and career readiness that they should be targeted for inclusion in new assessment systems." Linda Darling-Hammond, Joan Herman, James Pellegrino, Jamal Abedi, J. Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, Kenji Hakuta, Andrew Ho, Robert Lee Linn, P. David Pearson, James Popham, Lauren Resnick, Alan H. Schoenfeld, Richard Shavelson, Lorrie A. Shepard, Lee Shulman, and Claude M. Steele. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education; Center for Research on Student Standards and Testing, University of California at Los Angeles; and Learning Sciences Research Institute, University of Illinois at Chicago, p. 7. <https://edpolicy.stanford.edu/publications/pubs/847>
22. McQuillan, Phelps, & Stotsky, p. 46.
23. Linda Darling-Hammond, et al., pp. 16-18. <https://edpolicy.stanford.edu/publications/pubs/847>
24. For an in-depth discussion of these governance issues, see Peter Wood's excellent Introduction to *Drilling Through the Core*, <http://www.amazon.com/gp/product/0985208694>



PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

185 Devonshire Street, Suite 1101, Boston, MA 02110

T: 617.723.2277 | F: 617.723.1880

www.pioneerinstitute.org