



Measuring the Power of Learning.®

Research Report

ETS RR-16-13

Opt Out: An Examination of Issues

Randy E. Bennett

June 2016

Discover this journal online at
Wiley Online Library
wileyonlinelibrary.com

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Senior Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Director

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Anastassia Loukina
Research Scientist

John Mazzeo
Distinguished Presidential Appointee

Donald Powers
Managing Principal Research Scientist

Gautam Puhan
Principal Psychometrician

John Sabatini
Managing Principal Research Scientist

Elizabeth Stone
Research Scientist

Matthias von Davier
Senior Research Director

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Gontz
Senior Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Opt Out: An Examination of Issues

Randy E. Bennett

Educational Testing Service, Princeton, NJ

Media reports have recently given significant attention to the *opt-out* movement, an organized effort to refuse to take standardized tests. Although the narrative often told in early press accounts was of a viral grass-roots effort led by parents who object to state-mandated testing, the reality has turned out to be more complicated. Through a synthesis of news accounts, research studies, survey results, and state and federal education department documents, this paper examines the opt-out movement and some of the dynamics that appear to underlie it. Several topics are covered, including the movement's extent, the demographics of those participating in it, how much time students devote to tests, what other factors might be motivating the movement, and the level of public support for testing in general. The paper concludes with suggestions for how the assessment community might respond to the concerns raised by the movement and by the general public.

Keywords opt out; state testing; teacher evaluation; equity

doi:10.1002/ets2.12101

In the United States, 2015 was the year of the *opt out*, where many parents declined to allow their children to take state-mandated assessments. Aided in their efforts by such national, state, and local organizations as United Opt Out and Opt Out Oregon, websites were launched, media campaigns initiated, demonstrations conducted, and lobbying efforts undertaken. Many reasons are given for parents' actions but among the more common are the amount of instructional time lost to test preparation and administration, the educational irrelevance of "bubble tests," the difficulty of new standards and assessments, the pressure placed on students and educators to perform, and the belief that the Common Core State Standards (CCSS) and tests are instruments of corporate-driven reform directed at privatizing education (NYC Opt Out, n.d.; Opt Out Oregon, n.d.; Parent Guide to Common Core and High-Stakes Testing, n.d.; Robertson, 2015).

The level of test refusal, high degree of organization, and media coverage in turn motivated a notable political response. The amount of testing became a presidential campaign issue, splitting some of the prominent Republican candidates (Severns, 2015). Even President Obama weighed in with a call to place limits on state standardized testing, noting that, "Learning is about so much more than just filling in the right bubble" (Associated Press [AP], 2015b, para 3).

Congress took up the issue as part of its reauthorization of the No Child Left Behind Act (NCLB) of 2001. NCLB had dramatically increased the federal testing mandate, requiring states to use assessment as part of school evaluation. The act mandated 95% student participation in state assessment, giving federal and state education officials the option to financially punish districts that did not meet that standard (Harris & Fessenden, 2015). The law's reauthorization, titled the Every Student Succeeds Act, maintains the participation requirement but allows states and districts to decide how to incorporate failure to meet it into the accountability system (Klein, 2015b).

States, too, have reacted. According to the Council of Chief State School Officers (CCSSO), as of 2015, 39 states were exploring approaches to reducing testing time (Ujifusa, 2015a). Those approaches included time limits, cutting specific exams through legislation, handing responsibility to new state commissions, or working directly with local schools.

As to opt out itself, state law and policy vary significantly. Some states explicitly permit it, some forbid it, and others leave districts to act on their own volition (Aragon, Rowland, & Wixom, 2015).

Why does opt out matter? It matters because state assessments are the only comparable measures of performance at the building level. NAEP, the National Assessment of Educational Progress, does not report at that level, nor is it aligned with state content standards. Furthermore, state assessments are the only measures of building performance disaggregated by demographic group. Opt out can distort those results, preventing parents, educators, policymakers, and the public from understanding the extent to which schools are effectively educating all children.

Corresponding author: R. E. Bennett, E-mail: rbennett@ets.org

In the next sections, I examine the extent of support for opt out, the demographics of those participating in it, how much time students devote to tests, what other factors might be motivating the movement, and the level of public support for testing in general. I conclude by discussing suggestions for how the assessment community might respond to the concerns raised by the movement and by the general public.

How Strong Is Support for Opt Out?

In December 2015, the U.S. Department of Education (USDE) identified 13 states (less than 1/4 of jurisdictions) as having missed the 95% participation requirement in the 2014–2015 school year (Ujifusa, 2015b). A review of the notification letters sent to states, available state documentation, and news coverage suggests that the incidence of nonparticipation across states, districts, and grades was, in fact, highly variable.¹

For example, in California, the most populous state, the nonparticipation rate over all grades tested in English language arts (ELA) and math was only 3% (California Department of Education, 2015), in Idaho, it was 2% (Idaho State Department of Education, 2015), in Oregon under 4% (Oregon Department of Education, 2015), and in Connecticut about 4% (Connecticut State Department of Education [CSDE], 2015). For Washington State, the nonparticipation rates in Grades 3–8 were 2% for ELA and 3% for mathematics (Office of the Superintendent of Public Instruction [OSPI], 2015), while for Maine, the median rates were 5% and 6%, respectively (C. Tucker, personal communication, December 28, 2015).² Higher rates were evident for Colorado, where the medians across grades were 11% in ELA and 10% in mathematics, and in Rhode Island, which had rates of 12% and 10% in those same two subjects, respectively (Colorado Department of Education, 2015; Rhode Island Department of Education, 2015). In New York, however, the federal limit was dramatically exceeded, with 20% of eligible students in Grades 3–8 not testing (New York State Education Department [NYSED], 2015; Ujifusa, 2015c). That amount, approximately 200,000 students, was several times larger than the number from the previous year (Harris, 2015a; Harris & Fessenden, 2015).

With respect to districts, on New York's Long Island and in some upstate districts, most eligible students did not participate. In the Chateaugay, Rocky Point, and Onteora Central school districts, the refusal rates were 90%, 80%, and 66%, respectively (Harris & Fessenden, 2015). However, in the state's largest district, the New York City schools, a refusal rate of just 1.4% was reported (Harris, 2015a), quite consistent with the 1% average observed in the nation's 66 largest urban school systems (Council of the Great City Schools [CGCS], 2015). Outside New York State, individual districts proved to underlie several of the USDE citations, including for states with acceptable overall participation rates (e.g., California, Idaho, Illinois, Wisconsin) ("Tens of Thousands Skip State Tests in Illinois," 2016; Ujifusa, 2015b).

The occurrence of opt out also appears to have been notably greater in high school than at the lower school levels. In Washington State, for example, the 11th-grade rate was 49% in ELA and 53% in mathematics (OSPI, 2015), many times higher than the elementary and middle school rates of 2% and 3%, respectively. In Maine, the high school refusal rates were 39% in ELA and 40% in mathematics (C. Tucker, personal communication, December 28, 2015). High-school rates appear to have played a role in causing the USDE citations of Connecticut, Delaware, North Carolina, and Washington (CSDE, 2015; OSPI, 2015; Ujifusa, 2015b). Greater high school nonparticipation might be ascribed to students who, nearing the end of their secondary career, felt little obligation to participate in state assessment, especially given competition from more personally consequential college admissions and advanced-placement examinations.

In sum, the sources cited above suggest that significant levels of nonparticipation were restricted in 2015 to a minority of states and, except for New York, Colorado, and Rhode Island, to relatively small subsets of their eligible test-taking populations.

In addition to incidence rates, polls offer evidence relevant to support for the movement. A national sample collected by Education Next in spring 2015 suggests little public sympathy (Henderson, Peterson, & West, 2016). In that poll, only 25% of members of the public supported allowing parents to decide whether their children are tested, while 59% were against parental choice. Among parents specifically, 32% favored opt out, with 52% opposed. Finally, most teachers dismissed opt out (57%); only 32% gave it their support.

A somewhat more mixed result was found in the spring 2015 Phi Delta Kappa (PDK) and Gallup national poll (PDK & Gallup, 2015). In that survey, 44% of the public sample felt that parents should not be allowed to excuse their child from standardized testing, whereas 41% felt that parents should have that option. Among public-school parents, the sentiments were reversed, with 47% supporting the right to opt out and 40% disapproving of it. However, when public-school parents

were asked if they would excuse their own child, only 31% said that they would; 59% indicated that they would not opt them out.

The poll results cited above are consistent with a position statement on assessment issued by the National PTA (NPTA, 2016). In that communication, the parent–teacher association indicated its opposition to state and district policies that allow opt out, advising that all students participate in state assessment.

Who Is Opting Out?

Data suggest that parents who opt out represent a demographically particular population segment. In New York State, which had the highest overall refusal in number and percentage, students without a valid reason for missing the assessment were much more likely to be White and to come from districts with little or only average need for supplemental resources (Harris, 2015a; NYSED, 2015; Ujifusa, 2015c). Not surprisingly, these students were also much less likely to be economically disadvantaged and less likely to be English language learners. Colorado nonparticipants, too, were generally more likely to be White and less likely to be eligible for free and reduced price lunch (Colorado Department of Education, 2015). Finally, among Washington State 11th graders, economically disadvantaged students were also less likely to opt out (Parr & Teed, 2015).

Results from the PDK and Gallup (2015) national survey suggest that demographic differences in views toward opt out go beyond Washington, Colorado, and New York. In that poll, 44% of Whites supported allowing parents to excuse their students from testing and 41% were opposed to such exclusion. In stark contrast, only 28% of Blacks supported opt out and 57% were against it. The comparable figures for Hispanics were 35% supporting and 45% opposing. When asked if they would exclude their own child, the majority in each group would not, but the differences among groups were clearly evident: 21% of Blacks, 28% of Hispanics, and 34% Whites would exclude their own children from testing, whereas 75%, 65%, and 54%, respectively, would not opt them out.

Differences among population groups in attitudes toward testing more generally might help explain the demographic associations coming from New York, Colorado, and Washington, as well as from the PDK and Gallup poll (2015). In that poll, minority-group parents appeared more supportive of testing than White parents: 72% of Black parents and 61% of Hispanic parents considered test scores either very or somewhat important for measuring the effectiveness of their community schools, in contrast to 55% of White parents. In an earlier national survey conducted by AP and the National Opinion Research Center (NORC; Tompson, Benz, & Agiesta, 2013), parents from lower income households had more positive views toward testing than those from higher income households. A total of 85% of parents earning less than \$50,000 a year said that regular assessment was very important or extremely important, in contrast to the 73% earning \$50,000–\$100,000 and the 63% earning over \$100,000 per annum. Significantly more parents earning less than \$50,000 a year—79%—said that standardized tests measure the quality of education at a school somewhat well or very well, whereas only 66% of parents earning \$50,000–\$100,000 and 65% earning over \$100,000 had that same response.

Why do these demographic differences in attitude toward testing occur? One possibility is that they derive in part from the reality of the schools that different groups experience and how the use of state tests under NCLB might have affected that reality. Davidson, Reback, Rockoff, and Schwartz (2015) conducted a national analysis of school performance in the early years of the law's implementation. These investigators found that those institutions identified as failing to make adequate yearly progress 3 years running had a much higher percentage of students eligible for free and reduced-price lunch (a proxy for socioeconomic status [SES]) than did schools not identified in any of the same years (55% vs. 34%; Davidson et al., 2015, Table 1). In contrast to their nonproblematic counterparts, persistently failing schools on average served far lower percentages of White students (39.3% vs. 73.9%), and much higher percentages of Black (29.9% vs. 9.9%) and Hispanic (23.8% vs. 11.4%) students.

Not only are certain demographic groups more likely to experience failing schools but they are, not surprisingly, less likely to be served by high-quality teachers. Goldhaber, Lavery, and Theobald (2015) analyzed education data bases from Washington State. For the 2011–2012 school year, these investigators linked students to test scores and to their teachers in mathematics and reading courses in Grades 3 through 10. Teacher quality was indicated by experience, licensure exam score, and value-added estimates of effectiveness based on student test scores. Student disadvantage was measured by eligibility for free and reduced price lunch, being a member of an underrepresented minority group, and being in the lowest achievement test quartile in the prior grade. The investigators found that regardless of how teacher quality was measured or student disadvantage defined, teacher quality was inequitably distributed at virtually every school level.

The inequities documented in the two above studies are, of course, longstanding ones that NCLB was intended to reduce, in part by identifying which schools were failing to educate members of traditionally underserved groups effectively. For families in low-income areas who were experiencing persistently failing schools, the law required relief in the form of opportunity to transfer to nonfailing schools in the same district, school-supported private tutoring, and ultimately school restructuring or closure. For such families, state tests provided a very tangible benefit.³ In contrast, families whose children attended more successful schools might have seen little personal return from the state assessment and, sometimes, even loss due to the forced transfer of highly qualified teachers to failing schools as a means of achieving more equitable distribution (Shanklin, 2004).

Given the clear differences in attitudes about, participation in, and supposed benefit from state testing among racial/ethnic and socioeconomic groups, it should not be surprising that opt out has become a civil rights issue. On May 5, 2015, that issue was brought front and center with the release of a statement by 12 organizations (Leadership Conference on Civil and Human Rights, 2015). The organizations represented the African American, Hispanic American, disability, and female communities, and included the Leadership Conference on Civil and Human Rights, NAACP, National Urban League, National Council of La Raza, and the Disability Rights Education and Defense Fund. The statement said:

The educational outcomes for the children we represent are unacceptable by almost every measurement. And we rely on the consistent, accurate, and reliable data provided by annual statewide assessments to advocate for better lives and outcomes for our children. These data are critical for understanding whether and where there is equal opportunity . . . Until federal law insisted that our children be included in these assessments, schools would try to sweep disparities under the rug by sending our children home or to another room while other students took the test. Hiding the achievement gaps meant that schools would not have to allocate time, effort, and resources to close them. Our communities had to fight for this simple right to be counted and we are standing by it. . . . But we cannot fix what we cannot measure. And abolishing the tests or sabotaging the validity of their results only makes it harder to identify and fix the deep-seated problems in our schools. (The Leadership Conference, 2015, para 3, 5, 8)

The fears of distortion expressed by the Leadership Conference begin to become apparent in the data from New York, in which it was possible for state education officials to evaluate the relationship between opt out in 2015 and test scores from 2014—that is, before refusal had become a significant phenomenon. Although the above-reported demographic differences might suggest a positive relationship of scores to test refusal, students opting out were instead somewhat more likely *not* to have achieved proficiency on the prior year's exams when compared with the expected 2015 test-taking population (NYSED, 2015; Ujifusa, 2015c).⁴ A negative relationship to proficiency, but this time conditional on demographic group, was reported by Chingos (2015) in a secondary analysis of data from 648 of New York's 695 school districts. As might be expected from the individual student data, this district-level analysis found opt out to be less common in economically disadvantaged districts compared to the more affluent ones. After controlling for SES, districts with lower 2014 test scores had higher 2015 opt-out rates. Such a result might occur if, within SES strata, district staff encouraged opt outs in an effort to mask poor performance or parents pushed it to protect children who did not score well in 2014.⁵

How Much Time Do Students Spend Taking Tests?

One of the more common complaints voiced by opt-out advocates is that too much time is spent on testing, thereby detracting from learning and instruction (Boss, 2014; Satullo, 2015). Not surprisingly, one of the most common official responses to opt out has been to reduce—or otherwise limit—testing time. The Florida legislature passed a 45-hour annual limit, Texas eliminated 10 end-of-course exams, Virginia dropped exams in several grades, and New York has twice trimmed the length of its tests (AP, 2015a; Education Accountability Act, 2015; Harris, 2015b; Lazarin, 2014; Postal, 2015). Perhaps most prominently, the Obama administration called for a 2% cap on the percentage of instructional time devoted to state-mandated tests (USDE, 2015). But how much time do students actually devote, and who is requiring that time expenditure?

Several analyses offer an answer. CGCS conducted an inventory of the tests used by its 66 large urban district members (CGCS, 2015). Via an online survey and an analysis of district testing calendars, the study authors estimated the time devoted to state and district mandated testing for all students in the 2014–2015 school year. They found that the average student took eight standardized tests annually, including two NCLB-required ones and six formative/benchmark exams.⁶

Across the 66 member districts, an average of 1.9–2.3% of instructional time (~20–25 hours) went to state or district mandated tests in Grades 3–11.^{7,8}

CGCS also reported time estimates for other assessment types. On average, between 6.8 and 8.9 hours were devoted to PARCC/Smarter Balanced assessments, 8.5–10.8 hours to formative/benchmark assessments, 8.2–9.8 hours to student learning objectives assessments (primarily for students in NCLB nontested subjects and grades), and 7.5–9.3 hours for all other assessments that the state or district required for all students. Although these test-type averages are not directly comparable to the total averages (because they are calculated from only the districts that used them), the estimates do suggest that state-mandated NCLB assessments such as PARCC and Smarter Balanced might be responsible for something less than the majority of testing time.

A study by Lazarin (2014) appears to support that supposition. Lazarin used district and state assessment calendars, correspondence with school district and state central-office staff, and other publicly available information to identify the number and frequency of district and state-required standardized assessments and to determine the time it took for students to take the assessments. Included were only those tests that either the state or district required of all students. Findings showed that districts generally required more tests than states across all grade spans, with much of the district-level testing coming from interim benchmark exams. Students in NCLB grades took between 1.6 and 1.7 times more district-level exams than state exams, high school students took twice as many district as state exams, and K-2 students were given three times as many district as state assessments. Even so, students did not spend a great deal of time actually taking tests, including in the NCLB grades. On average, only about 1.6% of instructional time was devoted to standardized assessment.

Lazarin also discovered that district-level testing occurred more frequently and took more instructional time in urban than in suburban districts, a finding that might be related to the greater value ascribed to testing among civil rights groups and low-SES parents. In Grades K-2, urban students spent about 52% more time on district tests than their suburban peers, whereas in Grades 3–5 and 6–8, urban students spent approximately 80% and 73% more time, respectively, taking district-mandated examinations than did suburban students. The differential was greatest, however, for urban high school pupils, who devoted 266% more time to district-level exams than did their suburban counterparts (Lazarin, 2014).

A study by Teoh, Coggins, Guan, and Hiler (2014) supports Lazarin's findings with respect to both total time and urban/suburban differences. These investigators reviewed assessment calendars and guidelines, and communicated with administrators in 32 school districts, including large urban ones and some of their surrounding suburban counterparts. Teachers were also surveyed in six of the districts. Analyses concentrated on kindergarten, third, and seventh grades in the urban districts. Across the 12 urban districts, the average amount of time students spent on state *and* district tests was 1.7% of instructional time in third and seventh grades, and substantially less in kindergarten. In Grades 3 and 7, approximately 10 hours on average were devoted to state and district testing for the ELA and 7 hours for mathematics, making for 17 hours in total. Students in suburban districts, in contrast, devoted an average of 1.3% or less of instructional time to testing (or about 13 hours).⁹

Finally, a study by Guindon, Huffman, Socol, and Takahashi-Rial (2014) gives average time estimates reported from an online survey of North Carolina school districts taken in July 2014. Estimates were reported separately by grade. Over 99 districts, the median time devoted to state and local mandated assessment for Grades 3–11 was 16 hours, or 1.6% of instructional time.¹⁰ The highest estimate was from eighth grade students, who devoted 24 hours (2.3% of time). Although, in contrast to other studies, more time was spent on state than on district-required assessment, district assessment still accounted for a considerable portion of total testing time, about a third on average.

A different kind of evidence comes from the PDK and Gallup (2015) survey, which asked public-school parents the extent to which they agreed that their child complains about taking too many standardized tests. Consistent with the limited time that appears to be devoted to testing in the analyses above, most parents reported no such testing complaints—39% disagreed outright that their child complains and 24% responded neutrally, with only 31% indicating that their child had voiced concerns. And of that 31%, just 16% strongly agreed that they were hearing complaints.

As the research appears to indicate, the total time devoted to state and district assessments does not appear to be especially excessive on average, either in percentage terms or in hours. In hours, the estimates reported by Teoh et al. (2014), CGCS (2015), and Guindon et al. (2014) are only about a third to a half of the 45-hour cap signed into Florida law in 2015 for state and district-required assessments. In percentage terms, the reported averages for state *plus* district testing are generally similar to the Obama administration's recommendation for state testing *alone* (USDE, 2015). A second

notable result is that much (and in some studies, most) of the time students spend taking mandated tests comes from district-level requirements. Finally, suburban students, who are reported to have substantially higher opt-out rates than their urban counterparts, appear to spend less time in testing than city students.

It is important to note that the above analyses did not probe time spent on test preparation, a question about which there has been little recent survey research. A study by Rogers and Mirra (2014), which focused on learning time allocation generally, is a notable exception. These investigators received survey responses from 783 teachers sampled in fall 2013 to be representative of California's high schools. Results were reported for high-, mixed-, and low-poverty schools based on the percentage of students receiving free and reduced price lunch. Time figures were reported as a single estimate combining test preparation for district, charter, and state mandated tests, with administration of district and charter tests (but not state tests).

Similar to the urban/suburban differences reported above, time allocation varied as a function of school poverty level. In high-poverty schools, about 7.8 days per year on average were devoted to testing activities, which equates to approximately 47 instructional hours or 4.5% of instructional time. For low- and mixed-poverty schools, the allocations were considerably smaller: 4.4 and 4.7 days, respectively, translating to ~26 and ~28 instructional hours, or 2.5% and 2.7% of instructional time.¹¹ Notably, in all three types of schools, the instructional time lost from teacher absence, disrupted days (e.g., assemblies, safety lockdowns), and special days (e.g., prom, just before and after vacations and breaks) was almost twice the amount allocated to test preparation and testing (Rogers & Mirra, 2014, p. 14). Even adding in time for state test administration (excluded by Rogers and Mirra) would not approximate the size of the loss from these other sources.

What Is the Impetus Behind Opt Out?

Despite the fact that reducing testing time is a recurring political response, the evidence described thus far suggests that the actual time devoted to testing might not provide the strongest rationale for opting out, especially in the suburban low-poverty schools in which test refusal appears to occur more frequently. Another commonly expressed concern is that the high-stakes nature of tests makes children extremely anxious and often ill (e.g., Anand, 2013; Strauss, 2013), so perhaps it is this concern that is helping to motivate the movement. However, of 44 states surveyed, very few planned to use test scores in 2015 for making *any* student-level decisions (Mongeau, Felton, & Butrymowicz, 2015). Additionally, as noted above, most parents contacted by PDK and Gallup (2015) reported no complaints related to the frequency of testing from their children. Had testing been making significant numbers of students ill or extremely anxious, a single occasion ought to have been too much. If the evidence supports neither time nor anxiety as strong factual rationales, what might the impetus for opt out be?

NCLB mandated a variety of school accountability requirements for states in return for federal Title 1 funding to be directed at schools serving low-income students. Among the requirements were the implementation of academic content and performance standards, annual testing of students with respect to those standards, and the disaggregation of school-level scores by demographic group. These requirements were put into place because of concern that US elementary and secondary education was no longer internationally competitive and because of wide disparities in education quality and achievement for traditionally underserved groups.

The large variation among states in the quality of the academic content and performance standards that they subsequently implemented led in 2009 to the states launching, under the auspices of the National Governors Association and the CCSSO, an effort to create a uniform and considerably more rigorous set of content standards. Appearing the next year, the CCSS were adopted by 42 states, the District of Columbia, four US territories, and the Department of Defense Education Activity (Common Core State Standards Initiative [CCSSI], 2015). Contemporaneously, as part of the American Recovery and Reinvestment Act (ARRA) of 2009, the federal government offered grants through the Race to the Top Assessment Program for state consortia to develop assessments of common standards (though not necessarily of the CCSS).

Also as part of ARRA, grants were made available to individual states through the Race to the Top fund. Coming at the height of the Great Recession, the \$4.35 billion offered was very attractive to financially distressed states. The grants encouraged the adoption of common standards and common assessments. Of special note is that the grants gave preference to states that committed to institute more rigorous teacher-evaluation systems, systems sometimes perceived as intended to reduce the power of teacher unions, which were seen as roadblocks to education reform (Brill, 2010). Particular advantage was given to applicants that agreed to incorporate measures of students' academic growth on common state assessments and use those measures in compensation, promotion, tenure, and removal decisions (Lazarin, 2014; USDE, 2009).

The policy goals of ARRA were further advanced through USDE's NCLB waivers (CGCS, 2015). Beginning in 2011, the waivers allowed states to temporarily set aside some of the outcome accountability requirements of the law (including that calling for 100% of students to achieve proficiency by 2014; Center on Education Policy [CEP], 2012). In return, states committed to an ARRA-like regimen of adopting college-and-career-ready standards, assessments, and educator evaluation systems that used student growth as a significant contributor (CGCS, 2015).

Given the policy focus of ARRA and the waiver program, it is not surprising that, as of 2015, the overwhelming majority of states were either already requiring or transitioning to the use of test scores for some portion of teacher evaluation (Doherty & Jacobs, 2015; Mongeau et al., 2015). For subjects and grades not tested under NCLB, states had to find alternative measures, which motivated many districts to institute local pre- and posttests ("student learning objective" assessments) for calculating a student academic-growth measure.

It appears to have been the confluence of a revamped teacher evaluation system with a dramatically harder, Common Core-aligned test that galvanized the opt-out movement in New York State (Fairbanks, 2015; Harris & Fessenden, 2015; PBS Newshour, 2015). For 2014, 96% of the state's teachers had been rated as effective or highly effective, even though only 31% of students had achieved proficiency in ELA and only 36% in mathematics (NYSED, 2014; Taylor, 2015). These proficiency rates were very similar to ones achieved on the 2013 NAEP for Grades 4 and 8 (USDE, 2013a, 2013b, 2013c, 2013d). The rates were also remarkably lower than on New York's pre-Common-Core assessments. The new rates might be taken to imply that teachers were doing a less-than-adequate job and that supervisors, perhaps unwittingly, were giving them inflated evaluations for it.

That view appears to have been behind a March 2015 initiative from New York Governor Andrew Cuomo (Harris & Fessenden, 2015; Taylor, 2015). At his request, the legislature reduced the role of the principal's judgment, favored by teachers, and increased from 20% to 50% the role of test-score growth indicators in evaluation and tenure decisions (Rebora, 2015). As a result, the New York State United Teachers union urged parents to boycott the assessment so as to subvert the new teacher evaluations and disseminated information to guide parents specifically in that action (Gee, 2015; Karlin, 2015).

Not surprisingly, dislike for using standardized tests for teacher evaluation appears to go well beyond New York. The AP-NORC national survey conducted in 2012–2013, for example, found that 56% of parents who were teachers opposed using standardized test scores to judge teacher quality, compared with only 36% of other parents (Tompson et al., 2013).

More recent data suggest that parents' views on this topic might have become more negative since 2013 (though differences in respondent populations and question wording make comparisons tenuous). For example, PDK and Gallup (2015) reported that 63% of public-school parents and 55% of respondents generally were against using students' test performance in the evaluation of teachers, with only 37% and 43%, respectively, in favor.

What might underlie such negative views is suggested by Jeanette Deutermann, a parent from North Bellmore, New York, who stated the following: "The minute they tied teacher evaluations to those tests, they set up the classrooms to be about nothing except testing. ... So, of course, [teachers are] going to make kids spend all of their time preparing for the test. Their careers depend on it" (PDK & Gallup, 2015, p. K5).¹²

Ironically, significant segments of the educational research community might well side with the majority views reported by PDK and Gallup (2015). The research community has, in fact, consistently argued for caution in using students' test performance for teacher evaluation because other factors affect student achievement that are both beyond the control of classroom teachers *and* our ability to remove methodologically. Policy papers outlining such concerns have been published by the American Educational Research Association (AERA, 2015), AERA and the National Academy of Education (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2011), American Statistical Association (ASA, 2014), Economic Policy Institute (Shavelson et al., 2010), and individual researchers (e.g., Haertel, 2013). Given the New York experience, the research community's concerns, expressed at least as early as 2010, were perceptive. Those concerns remain important given that many states with existing evaluation systems are still relatively early in the process of introducing new curriculum, new assessments, and new performance standards, as earlier done to such negative effect in New York.

The chances of repeating that experience may be reduced substantially, however, by a changing political climate. At the federal level, USDE issued waivers in 2015 allowing some states to defer their use of test results for teacher evaluation (Klein, 2015c). That action was followed by passage of the Every Student Succeeds Act, which removes such federal evaluation requirements, superseding the waivers as of August 2016 (Klein, 2015a, 2015b).¹³ Passage of the act has also changed the enforcement focus of USDE. Some states, having earlier agreed to federal evaluation requirements, never implemented

or subsequently stepped back from them (Doherty & Jacobs, 2015; Klein, 2015d), violations that USDE no longer appears likely to pursue (Klein, 2015a). Other states, however, codified those requirements in law during the Race to the Top fund competition, thereby institutionalizing test-based educator evaluation practices. Those laws or their implementation may begin to shift with the results of legal action and new thinking among policymakers. A significant number of lawsuits challenging evaluation practices have been filed against states and districts by unions and their affiliates (Education Week, 2015). In addition, CCSSO has released principles that encourage the use of multiple evaluation measures (lessening the dependence on test scores), and that move the purpose toward evaluation for teacher support (decreasing attention to high-stakes uses) (CCSSO, 2016).

How Does the Public View Testing?

The significant media attention given to the opt-out movement, as well as the broader controversy surrounding the Common Core, should raise questions about the extent of public support for educational assessment in general. Recent data on this question come from three surveys, two of which were conducted in spring 2015 using nationally representative samples. From the Education Next survey, Henderson et al. (2016) reported that 67% of respondents favored the federal requirement for annual testing, 21% were opposed, and the remainders were neutral. Support among parents was about as high as that of the public as a whole. Teachers, however, were divided, with 47% in favor and 46% against continuing the policy for federally mandated testing.

In the PDK and Gallup (2015) poll, 67% of respondents and 71% of public-school parents felt that using tests to measure what students have learned was either very or somewhat important for improving public schools in their community. Similarly, 57% of respondents and 56% of public-school parents felt that the scores students receive on standardized tests were either very or somewhat important to measuring the effectiveness of their community's public schools.

At the same time, 64% of respondents overall and 67% of public-school parents felt that there was too much emphasis on testing in the public schools in their community, as compared with 26% and 28% who felt that the emphasis was either about right or not enough (PDK & Gallup, 2015). Furthermore, when compared with other selected indicators, tests were the least preferred (PDK & Gallup, 2015). Lower percentages of respondents and of public-school parents felt that the most accurate picture of a public-school student's academic progress would be provided by standardized test scores (16%), as opposed to teacher grades (21% of respondents and 22% of public-school parents), teacher's written observations (26% and 25%), and examples of the student's work (38% and 37%).

The last survey was conducted online in August 2014 by CGCS (2015). The survey used a sample of 600 parents whose children attended Great City districts implementing the CCSS. While the survey's online collection method, focus on large urban districts, and restriction to those implementing the CCSS limit generalizability, the findings are notable for their consistency. Results indicated strong support for "measurement" and its role in accountability but also for what constitutes a "better test." Large majorities of respondents agreed with the following statements: It is important to have an accurate measure of what my child knows (83%); Accountability for how well my child is being educated is important, and it begins with accurate measurement of what he or she is learning in school (78%); Better tests would expect students to demonstrate their thinking, thinking critically, and solving complex problems (71%); Better tests would ask students to do more than provide an answer by filling in bubbles or picking multiple choice answers (69%); and Children should be required to take tests that ensure they're learning the standards, new tests should replace current tests (67%).

In combination, these results suggest that the public may have more favorable views toward testing than either the existence of the opt-out movement or the extensive media coverage given it would imply. But the results also indicate a perception that tests as currently constituted could be markedly improved and that they are neither the only assessment method nor necessarily the best available approach.

Discussion

In this paper, I examined the opt-out movement and the dynamics that appear to be behind it, factors critical to understand in formulating action recommendations. Opt out matters because state assessments are the only comparable measures of building-level performance within a state and the only building-level measures disaggregated by demographic group. To the extent that they adequately reflect state standards, these assessments can give education officials information for localized action and advocates a basis for getting resources directed at underperforming, low-SES schools.

According to USDE, 13 states failed to meet federal assessment participation requirements for the 2014–2015 school year. Within and across these states, however, the incidence of nonparticipation was highly variable, with 20% of eligible third through eighth graders in New York State not testing in 2014–2015, but only 3% not participating over all assessed grades in California. National polls suggest that the general public opposes opt out, though the margins vary considerably between polls. Moreover, polls suggest that by and large the public supports the use of standardized tests for measuring school effectiveness, even if it feels such tests might not be the best method.

Parents who opt their children out appear to represent a distinct subpopulation. In New York, opt outs were more likely to be White and *not* to have achieved proficiency on the previous year's state examinations. Those students were less likely to be economically disadvantaged, to come from districts serving relatively large numbers of poor students, and to be an English language learner. Similar associations for race and SES occurred in Colorado and Washington. These demographic associations are consistent with attitudes toward testing, which polls suggest is perceived less favorably by Whites and higher-income cohorts than by members of minority and lower-SES groups. These differences in perception and action have led to opt out becoming a civil-rights issue since it has the potential to distort state test results, complicating the identification of schools and districts that are failing to educate traditionally underserved students effectively.

Although a frequent rationale for opt out has been the time devoted to mandated tests, studies suggest that only about 2% of instructional time on average is used for test administration and much of that time is associated with district-required measures. The very limited research on test preparation suggests that, at least in high school, such activities would take only an additional percentage point or two of instructional time and, in any case, considerably less time than given to common sources of instructional distraction.

A more powerful motivator, at least in New York State, appears to have been a dramatic increase in the role of student test results for teacher evaluation, a use with both limited support among the public nationally and in the educational research community. While the mechanisms leading from that increase to opt out have not been systematically documented, news reports suggest that the trigger was the combination of this controversial test use with a (Common-Core-aligned) assessment on which notably lower percentages of students were expected to achieve proficiency. That combination likely encouraged the proliferation of such questionable educational practices as narrow and excessive test preparation and unreasonable pressure on students, resulting in significant discord among educators, parents, and children. That discord, in turn, led educators and a particular demographic of parents to mobilize, local and national media to respond, the movement to spread (taking on members with a wider variety of concerns), and politicians to react.

Why did other states, including those cited by USDE, appear to have considerably lower nonparticipation rates? Some states, like California, did not link teacher evaluation to student test scores at all (Doherty & Jacobs, 2015). Other states made such linkages but did not make test scores the preponderant evaluation criterion (e.g., Idaho, Maine); got USDE permission to delay implementation (e.g., Connecticut, Delaware, Idaho); or stepped back from their original policies altogether (e.g., Wisconsin; Doherty & Jacobs, 2015; Klein, 2015c).¹⁴ Finally, most states avoided direct confrontation with teacher unions.

The summary above should make clear that opt out is a complicated, politically charged issue made more so by its social class and racial/ethnic associations. It is also an issue that appears to be as much about test use as about tests themselves. While the majority of the public opposes opt out, the minority that supports it is sizable, organized, vocal, and politically effective. Given these observations, how should the assessment community respond to the concerns raised by the movement as well as to the ones voiced by the general public?

The community might best respond through two categories of action. The first category is more active and effective communication targeted at policymakers, state department staff, local educators, parents, students, and the public. Mechanisms that might be tried include the same public-relations vehicles used by the opt-out movement to such considerable effect— websites, radio and television announcements, and media interviews. Such a campaign might be launched by the community's professional organizations to foster greater understanding of the value of *high-quality* assessment and its appropriate use.

What particular messages should this campaign communicate? One message is that the competencies students need for success in college and careers are changing and their levels are increasing (Organisation for Economic Co-operation and Development, 2012, p. 6). Students who develop these (cognitive and noncognitive) competencies are likely to have a better chance at sustained, productive employment than those who do not. As a consequence, any measure that we use to evaluate the effectiveness of our schools in imparting these competencies will necessarily be broader and more demanding

than past measures, whether those measures are linked to the CCSS or to standards being newly promulgated by some states individually.

A second message is that we agree with opt-out advocates, President Obama, and parent surveys that voice concerns about the limited relevance—and negative instructional effects—of relying too heavily on multiple-choice tests (Fredriksen, 1984). A greater diversity of task types, including simulations and performances, will be needed if we are to know whether students are on track to master needed competencies. Can they write using evidence from sources, read in digital environments, collaborate in solving science problems, and mathematically model problem situations? All of these competencies are characteristic today of the activities undertaken in advanced academic settings and the workplace. All are also difficult to measure with traditional tests. NAEP and the Common Core assessments have made important moves in these directions such that the term *standardized test* no longer needs to mean “fill in the bubbles.”

Third is that such assessments will take significant class time if they are to represent meaningfully what students know and can do. When assessments are too short, they are not able to cover the breadth and depth of the standards effectively. Reducing test length may, in fact, have the unintended effect of encouraging teachers to pursue even narrower test preparation activities as they seek to instruct to the more limited content of the shortened assessment, thereby further incenting parents to opt their children out. While advances like adaptive testing can help reduce test length, the content constraints required for standards-based assessment and the fact that adaptive testing cannot be straightforwardly applied to performance tasks suggest that the time savings may be limited. In brief, discussion might better focus on whether the time spent is justified by the quality of the measure and what is lost when time is reduced.

A fourth message is that the use of student test scores to evaluate teachers is a highly controversial practice, even within the educational research community. Evidence supporting the use of such scores for evaluating student achievement does not make those scores automatically valid for decisions of compensation, promotion, tenure, and removal, even if scores are used as only one of several indicators. That said, the need to evaluate educator performance is a legitimate one and student test results might play some role, pending validation evidence. To facilitate buy-in, any such evaluation systems might best be constructed as part of a cooperative effort among teachers, principals, parents, unions, state officials, and assessment community members, rather than being dictated by policymakers alone.

Fifth, participation is essential if student competency and educational effectiveness are to be evaluated fairly (The Editorial Board, 2015). Selective participation, a consequence of opt out, distorts results, making it harder for policymakers to direct attention and resources to the districts, schools, and students that need them. Given that selective participation is associated with demography, opt out may undermine legal mandates to monitor educational opportunities for traditionally underserved groups and compromise the quality of education offered them.

The second category of action is to translate words into deeds. It is clear that significant segments of the public, the education community, and policymakers do not want to spend more student time on the types of tests with which they are familiar (i.e., multiple-choice assessments perceived as distant from the format and content of instruction, and therefore irrelevant for it). That sentiment is apparent in the material on opt-out websites, in the public's responses to surveys, and in President Obama's remarks (AP, 2015a; CGCS, 2015; NYC Opt Out, n.d.; PDK & Gallup, 2015).

One challenge will be to create tests that offer more actionable results. Because of their survey nature, state summative assessments will necessarily be limited in this regard. However, if designed, for example, in keeping with theory-based learning progressions, such tests should at the least be able to give teachers a starting point for formative follow-up (Bennett, 2011, p. 7; Bennett, Deane, & van Rijn, 2016). Similarly, the analysis of student solution processes in completing performance tasks might offer teachers potential directions for strategy development (Bennett, Persky, Weiss, & Jenkins, 2010).

A second, perhaps more difficult challenge, will be to devise tests that look like—and are—valuable learning experiences in addition to measuring devices. The community needs to build assessments that encourage participation because parents, teachers, and students see participation, *and preparation*, as worthwhile learning endeavors.

Reconceptualizing test preparation as worthwhile becomes easier to the extent that test content and format themselves more clearly represent the depth and breadth of the standards, as opposed to the highly restricted subsets of it that multiple-choice tests are often perceived to embody by opt-out advocates, the public, and policymakers. Under that former scenario, preparation for the test should be better aligned with what instruction would have been in the test's absence.

Even when tests more adequately represent the depth and breadth of content standards, however, that representation will not be complete simply because of practical limits on testing time and the types of tasks that can be presented in

a standardized assessment. As a consequence, excessive preparation that focuses only on those task types and on the competencies most likely to be assessed can be problematic, leading to curricular narrowing and score inflation (Koretz, 2010; Koretz & Hamilton, 2006). Minimizing the role of test scores in educator evaluation and encouraging a concentration throughout the year on teaching to the standards should help dampen these potential negative effects.

It is important to note that, as long as it is not excessive, taking practice tests aligned to standards can be a legitimate part of instruction. Deliberate practice helps learners develop fluency for basic procedures, acquire qualitative understanding, consolidate knowledge, and connect it to conditions of use (Ericsson, Krampe, & Tesch-Römer, 1993; National Research Council, 2000, p. 125; VanLehn & van der Sande, 2009). Taking a well-designed practice test can, itself, help promote knowledge retention and transfer and do so more effectively than common approaches to studying (Butler, 2010; Hinze, Wiley, & Pellegrino, 2013; Paul, 2015; Roediger & Karpicke, 2006; Rohrer & Pashler, 2010).

While test design can make preparation more valuable, design can also make participation worthwhile. Several mechanisms for modeling good teaching and learning practice are described by Bennett (2015, pp. 382–386). Among these mechanisms are including knowledge representations in the assessment that are similar to the ones proficient performers use in their domain practice (e.g., planning tools for writing, standards for what constitutes quality work), and structuring at least some performance tasks so that they recapitulate the sequence of steps that might be found in an extended project. Such knowledge representations and decomposed task structures, if repeatedly encountered on summative and formative assessments, should then be more likely to become routine parts of teaching and learning practice.

Assessment design can also be oriented toward increasing student engagement. Potentially productive directions include the incorporation of interactive elements and animation in summative assessment, as well as the building of formative assessments into educational games. The former might make summative tests appear more game-like and the latter create more positive associations for students with assessment. More questionable in the context of opt out is the idea of substituting student interactions incidentally gathered from electronic learning environments for a summative test. The constant recording of student behavior, the introduction of consequences (like teacher evaluation) into learning interactions, and the highly contextualized nature of results raise privacy, instructional efficacy, and validity concerns that may create more issues than they resolve (see Bennett, 2015, pp. 391–395, for a more complete discussion).

The additional testing requirements that have been levied by many school districts underscore the need for more coherent systems of assessment (Bennett & Gitomer, 2009; Pellegrino, Chudowsky, & Glaser, 2001). In a coherent system, each assessment has a clear purpose and the different assessments—summative and formative—work together to facilitate teaching and learning. Efficiency, clarity of purpose, synergy, and utility are likely to be higher when the testing program is purposefully designed instead of assembled more incidentally. To help districts rationalize their existing systems, Achieve, an educational nonprofit, published the Student Assessment Inventory for School Districts (Achieve, 2014). With similar intent, New York State education officials reviewed each district's assessment portfolio, issued recommendations for eliminating local assessments, and posted the recommendation letter on the state website (Lazarin, 2014). In addition, the state awarded 31 competitive grants to districts as incentives for local officials to review their district testing programs. That these costly actions were taken only serves to highlight the incoherence that may characterize many local systems. As consultants, assessment community members can help state and local officials enhance coherence through better articulation and reduced duplication among their test offerings.

Finally, a theory of action should be explicated for state, as well as district, assessment programs (Bennett, 2010). That theory should describe the assessment program's intended effects, and precisely how it proposes to achieve them, in ways that all stakeholders can comprehend. The theory can then become the basis for examining coherence, communicating the intended value of the testing program, and getting early indications as to its political viability and scientific defensibility.

The suggestions posed above, and the preceding conclusions drawn about the nature of opt out, were based on an examination of selected news accounts, survey results, research studies, and state and federal education department documents. Sources were found through Internet searches and by following citations in located documents. Although every attempt was made to be balanced in the selection and interpretation of documents, the inferences made and the suggestions offered are unavoidably subjective. That subjectivity is a significant limitation of this paper, common to policy reviews of this type. Further research will be needed to determine whether the particular suggestions described would be likely to change opt-out behavior or affect public opinion in the ways intended.

A second limitation of this paper is that I did not explore the measurement implications of opt out. Opt out might be expected to distort equating, growth modeling, and trend analysis, among other things. Such effects should also be the subject of research as they are critical for the assessment community and for policymakers to understand.

In conclusion, this examination suggests that the assessment community might do well to become an active participant in the opt-out conversation. The incentives for the community to, itself, opt out are unfortunately strong. Participation is likely to be perceived as self-interested and responding to attacks on testing may well call more attention to them. That said, engaging critics to better understand their concerns, communicating with constituents to promote appropriate assessment practices, and creating approaches that have positive impact on teaching and learning may be the most constructive response that the community can offer.

Notes

- 1 I use *nonparticipation* rather than *opt out* because most states do not differentiate the reasons students failed to test.
- 2 I computed all percentages for Washington State from raw data downloaded from the Office of Superintendent for Public Instruction's website.
- 3 For a variety of reasons, relatively few eligible students appear to have taken advantage of NCLB's school transfer or tutoring provisions (Vernez et al., 2009). However, even if not acted upon, knowing one had those options would appear to be a benefit.
- 4 It might appear counterintuitive for New York State opt outs to be on average both more economically advantaged and lower scoring than the expected test-taking population. Because the student-level correlation between SES and achievement is only about .29 (Sirin, 2005), that result is not unreasonable.
- 5 The suggested tendency toward opt out among students who are both less proficient and, for example, White might appear to undermine the civil rights' concern because the result would appear to *increase* the size of measured achievement gaps, providing a stronger, rather than weaker, argument for the provision of resources to underserved groups. However, the essential point is that data that are not missing at random cause distortion, making it impossible to know whether the gaps are getting bigger or smaller and, therefore, whether schools are succeeding in closing them. As the Leadership Conference on Civil and Human Rights (2015, para 8) said, "... we cannot fix what we cannot measure."
- 6 *Formative assessment* and *benchmark assessment* were confounded in the survey questions (see CGCS, 2015, p. 149).
- 7 I calculated the 1.9% value from Footnote 5 and Table 3 in CGCS (2015, p. 28). Percentages appear to be based on an instructional year of 1,080 hours.
- 8 These values might overestimate actual time somewhat because the survey employed ranges and the maximum of each range was used for analysis (see Item 5, CGCS, 2015, p. 22).
- 9 The "13-hour" figure is my estimate, as Teoh et al. (2014) did not report how many hours they used as the baseline for total instructional time. Estimating from the reported percentages and hours, it appears that total instructional time equals 1,000 hours per school year.
- 10 I calculated this median estimate from data provided in Figure 9 of Guindon et al. (2014, p. 8).
- 11 I calculated percentages of instructional time based on an instructional day of 360 minutes and a year of 175 instructional days (1,050 hours), as indicated by Rogers and Mirra (2014, p. 7).
- 12 How much time teachers actually did spend on test preparation in New York does not appear to have been studied. If we consider the fact that California has no state policy linking teacher evaluation to student test results (Doherty & Jacobs, 2015), we might expect test preparation levels in New York to be greater than the ones implied for California high schools by Rogers and Mirra (2014).
- 13 See, for example, Section 1111(e)(1)(B)(iii)(IX) of the Every Student Succeeds Act.
- 14 The New York State Board of Regents announced in December 2015 that it too would suspend the use of test scores in teacher evaluation for a 4-year period (Moody, 2015).

References

- Achieve. (2014). *Student assessment inventory for school districts*. Retrieved from <http://www.achieve.org/files/AchieveStudentAssessmentInventory.pdf>
- American Educational Research Association. (2015). *AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs*. Retrieved from <http://edr.sagepub.com/content/early/2015/11/10/0013189X15618385.full.pdf+html>

- American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Retrieved from http://www.amstat.org/policy/pdfs/asa_vam_statement.pdf
- Anand, A. (2013, April 19). At Common Core talk, a principal says his reality includes vomit. *Chalkbeat New York*. Retrieved from <http://ny.chalkbeat.org/2013/04/19/at-common-core-talk-a-principal-says-his-reality-includes-vomit/#.VenW69JVhBc>
- Aragon, S., Rowland, J., & Wixom, M. A. (2015). *Assessment opt-out policies: State responses to parent pushback*. Retrieved from the Education Commission of the States website: <http://www.ecs.org/clearinghouse/01/17/68/11768.pdf>
- Associated Press. (2015a, October 19). *STAAR tests to get less time-consuming in grades 3 to 8*. Retrieved from <http://www.ksl.com/?nid=157&sid=36975952>
- Associated Press. (2015b, October 24). *Obama calls for capping time devoted to standardized tests*. Retrieved from <http://www.pbs.org/newshour/rundown/obama-calls-cap-class-time-devoted-standardized-tests/>
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8, 70–91. <http://dx.doi.org/10.1080/15366367.2010.508686>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice* 18, 5–25. <http://dx.doi.org/10.1080/0969594X.2010.513678>
- Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, 39, 370–407. <http://dx.doi.org/10.3102/0091732X14554179>
- Bennett, R. E., Deane, P., & van Rijn, P.W. (2016). From cognitive-domain theory to assessment practice. *Educational Psychologist*. Advance online publication. <http://dx.doi.org/10.1080/00461520.2016.1141683>
- Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment, and professional support. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61). New York, NY: Springer.
- Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for NAEP. *The Journal of Technology, Learning and Assessment*, 8(8). Retrieved from <http://files.eric.ed.gov/fulltext/EJ887134.pdf>
- Boss, C. (2014, November 2). Common Core stirs fear of too many tests. *Columbus Dispatch*. Retrieved from <http://www.dispatch.com/content/stories/local/2014/11/02/common-core-stirs-fear-of-too-many-tests.html>
- Brill, S. (2010, May 17). The teachers' unions' last stand. *The New York Times*. Retrieved from <http://www.nytimes.com/2010/05/23/magazine/23Race-t.html?pagewanted=all>
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 36, 1118–1133. <http://dx.doi.org/10.1037/a0019902>
- California Department of Education. (2015, December 22). *State schools chief Tom Torlakson announces 97 percent participation rate for Smarter Balanced assessments* (Release No. 15-91). Retrieved from <http://www.cde.ca.gov/nr/ne/yr15/yr15rel91.asp>
- Center on Education Policy. (2012). *Frequently asked questions regarding the Secretary of Education's waivers of major ESEA requirements*. Retrieved from http://www.cep-dc.org/cfcontent_file.cfm?Attachment=Riddle_FAQsonESEAWaivers_2.17.12.pdf
- Chingos, M. M. (2015). *Who opts out of state tests?* (Brookings Brown Center Chalkboard No. 115). Retrieved from <http://www.brookings.edu/research/papers/2015/06/18-chalkboard-who-opts-out-chingos>
- Colorado Department of Education. (2015). *CMAS PARCC 2015 participation results*. Retrieved from <http://www.cde.state.co.us/assessment/cmas-elamath2015-participation>
- Common Core State Standards Initiative. (2015). *Standards in your state*. Retrieved from <http://www.corestandards.org/standards-in-your-state/>
- Connecticut State Department of Education. (2015, September 2). *Update on student performance: First analysis of Smarter Balanced results 2014-15* [PowerPoint presentation]. Retrieved from <http://www.sde.ct.gov/sde/lib/sde/powerpointpresentations/evalresearch/connecticutassessmentresults201415sbe.pptx>
- Council of Chief State School Officers. (2016). *Principles for teacher support and evaluation systems*. Retrieved from <http://www.ccsso.org/Documents/2016/Principles%20for%20Teacher%20Support%20and%20Evaluation%20Systems.pdf>
- Council of the Great City Schools. (2015). *Student testing in America's great city schools: An inventory and preliminary analysis*. Retrieved from <http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Testing%20Report.pdf>
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. H., & Rothstein, J. (2011). *Getting teacher evaluation right: A background paper for policy makers*. Retrieved from [http://www.aera.net/Portals/38/docs/News_Media/AERABriefings/Hill%20Brief%20-%20Teacher%20Eval%202011/GettingTeacherEvaluationRightBackgroundPaper\(1\).pdf](http://www.aera.net/Portals/38/docs/News_Media/AERABriefings/Hill%20Brief%20-%20Teacher%20Eval%202011/GettingTeacherEvaluationRightBackgroundPaper(1).pdf)
- Davidson, E., Reback, R., Rockoff, J., & Schwartz, H. L. (2015). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB. *Educational Researcher*, 44, 347–358. <http://dx.doi.org/10.3102/0013189X15601426>
- Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning*. National Center on Teacher Quality. Retrieved from the National Center on Teacher Quality website: <http://www.nctq.org/dmsView/StateofStates2015>
- Education Accountability Act. Fla. Stat. § 2015-6. (2015). Retrieved from <https://www.flsenate.gov/Committees/BillSummaries/2015/html/1087>

- Education Week. (2015, October 6). *Teacher evaluation heads to the courts*. Retrieved from <http://www.edweek.org/ew/section/multimedia/teacher-evaluation-heads-to-the-courts.html?cmp=eml-enl-eu-news2>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406. <http://dx.doi.org/10.1037/0033-295X.100.3.363>
- Fairbanks, A. M. (2015, January 8). Teachers to the test. *The Atlantic*. Retrieved from <http://www.theatlantic.com/education/archive/2015/01/teachers-to-the-test/384306/>
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193–202. <http://dx.doi.org/10.1037/0003-066X.39.3.193>
- Gee, D. J. (2015, March 30). NYSUT president urges parents to opt out of state tests. *Buffalo News*. Retrieved from <http://schoolzone.buffalonews.com/2015/03/30/nysut-president-urges-parents-to-opt-out-of-state-tests/>
- Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44, 293–307. <http://dx.doi.org/10.3102/0013189X15592622>
- Guindon, M., Huffman, H., Socol, A. R., & Takahashi-Rial, S. (2014). *How much testing is taking place in North Carolina schools at grades K-12? An analysis of federal, state, and local required assessments*. Retrieved from the North Carolina Department of Public Instruction website: <http://www.dpi.state.nc.us/docs/intern-research/reports/testing2014.pdf>
- Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Princeton, NJ: Educational Testing Service.
- Harris, E. A. (2015a, August 12). 20% of New York State students opted out of standardized tests this year. *New York Times*. Retrieved from http://www.nytimes.com/2015/08/13/nyregion/new-york-state-students-standardized-tests.html?_r=0
- Harris, E. A. (2015b, September 16). New York will trim Common Core exams after many students skipped them. *New York Times*. Retrieved from http://www.nytimes.com/2015/09/17/nyregion/new-york-will-trim-common-core-exams-after-many-students-skipped-them.html?_r=1
- Harris, E. A., & Fessenden, F. (2015, May 20). ‘Opt-out’ becomes anti-test rallying cry in New York State. *New York Times*. Retrieved from <http://nyti.ms/1c4ZMBp>
- Henderson, M. B., Peterson, P. E., & West, M. R. (2016). The 2015 EdNext Poll on school reform. *Education Next*, 16(1). Retrieved from <http://educationnext.org/2015-ednext-poll-school-reform-opt-out-common-core-unions/>
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, 69, 151–164. <http://dx.doi.org/10.1016/j.jml.2013.03.002>
- Idaho State Department of Education. (2015). *Idaho report card: All grades*. Retrieved from <https://apps.sde.idaho.gov/ReportCard/SchoolYear/21>
- Karlin, R. (2015, March 30). Going on the offensive, NYSUT’s Magee calls for test boycott. *Timesunion*. Retrieved from <http://blog.timesunion.com/capitol/archives/231494/going-on-the-offense-nysuts-magee-calls-for-test-boycott/>
- Klein, A. (2015a, December 18). Ed. Dept. sketches out transition to ESSA from NCLB, previews regulation. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2015/12/ed_dept_sketches_out_transitio.html
- Klein, A. (2015b, November 30). ESEA reauthorization: The Every Student Succeeds Act explained. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2015/11/esea_reauthorization_the_every.html
- Klein, A. (2015c, September 15). Teacher-evaluation reins loosen under NCLB waivers. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2015/09/16/teacher-evaluation-reins-loosen-under-nclb-waivers.html>
- Klein, A. (2015d, November 4). Will teacher evaluations through test scores outlast Obama. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2015/11/will_teacher_evaluations_throu.html?cmp=eml-enl-eu-news2
- Koretz, D. (2010). *Implications of current policy for educational measurement*. Retrieved from the Educational Testing Service website: <http://www.ets.org/Media/Research/pdf/KoretzPresenterSession3.pdf>
- Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). Westport, CT: American Council on Education/Praeger.
- Lazarin, M. (2014). *Testing overload in America’s schools*. Retrieved from the Center for American Progress website: <https://cdn.americanprogress.org/wp-content/uploads/2014/10/LazarinOvertestingReport.pdf>
- Leadership Conference on Civil and Human Rights. (2015, May 5). *Civil rights groups: “We oppose anti-testing efforts* [Press release]” Retrieved from <http://www.civilrights.org/press/2015/anti-testing-efforts.html>
- Mongeau, L., Felton, E., & Butrymowicz, S. (2015, May 6). Survey: Impact of so-called ‘high stakes’ tests actually low. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2015/05/06/survey-impact-of-so-called-high-stakes-tests.html>
- Moody, R. (2015, December 15). Four years of relief on teacher evaluations after Board of Regents action. *Watertown Daily Times*. Retrieved from <http://www.watertowndailytimes.com/news03/four-years-of-relief-on-teacher-evaluations-after-board-of-regents-action-20151215>
- National PTA. (2016). *Position statement—assessment*. Retrieved from <http://www.pta.org/advocacy/content.cfm?ItemNumber=4716&navItemNumber=4616>

- National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: The National Academies Press.
- New York State Education Department. (2014, August 14). *State education department releases grades 3-8 assessment results* [Press release]. Retrieved from <http://www.nysed.gov/news/2015/state-education-department-releases-grades-3-8-assessment-results>
- New York State Education Department. (2015, August). *Measuring student progress in grades 3-8 English language arts and mathematics* [Press release]. Retrieved from <http://www.p12.nysed.gov/irs/pressRelease/2015800/documents/20153-8TestResultsCommunicationsDeck.pdf>
- NYC Opt Out. (n.d.). *Talking points*. Retrieved from <http://www.optoutnyc.com/test-info/talking-points/>
- Office of Superintendent of Public Instruction. (2015). Washington state report card: SBA scores by state [Data file]. Retrieved from <http://reportcard.ospi.k12.wa.us/DataDownload.aspx>
- Opt Out Oregon. (n.d.). *Why opt out?* Retrieved from <http://optoutoregon.org/whyoptout/>
- Oregon Department of Education. (2015). *Statewide report card: 2014-15*. Retrieved from <http://www.ode.state.or.us/data/annreportcard/rptcard2015.pdf>
- Organisation for Economic Co-operation and Development. (2012). *Better skills, better jobs, better lives: Highlights of the OECD skills strategy*. Retrieved from <https://skills.oecd.org/documents/SkillsHighlightsEnglish.pdf>
- Parent guide to Common Core and high-stakes testing* [Brochure]. (n.d.). Retrieved from <http://unitedoptout.com/common-core-high-stakes-testing-brochure/>
- Parr, A., & Teed, P. (2015, September 10). *2014-15 assessment results* [Slide presentation]. Retrieved from the Washington State Board of Education website: http://www.sbe.wa.gov/documents/BoardMeetings/2015/Sept/SBA_ResultsPresentation.pdf
- Paul, A. M. (2015). Researchers find that frequent tests can boost learning. *Scientific American*, 313(2), 1–7.
- PBS Newshour. (2015, May 26). *What galvanized standardized testing's opt-out movement*. Retrieved from <http://www.pbs.org/newshour/bb/galvanized-standardized-testings-opt-movement/>
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Phi Delta Kappa & Gallup. (2015). The 47th PDK/Gallup Poll of the public's attitudes toward the public schools: Testing doesn't measure up for Americans. *Phi Delta Kappan*, 97(1). Retrieved from http://pdkpoll2015.pdkintl.org/wp-content/uploads/2015/10/pdkpoll47_2015.pdf
- Postal, L. (2015, April 14). Gov. Scott signs bill to scale back testing in Florida public schools. *Orlando Sentinel*. Retrieved from <http://www.orlandosentinel.com/features/education/school-zone/os-scott-testing-bill-schools-post.html>
- Rebora, A. (2015, September 17). N.Y. regents question evaluation-scoring model, offer flexibility. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/teacherbeat/2015/09/ny_regents_approve_evaluation_.html?cmp=e-ml-enl-eu-news3
- Rhode Island Department of Education. (2015). *Rhode Island's PARCC results for students in grades 3 through 8 and high school—2015: A look into teaching and learning*. Retrieved from <http://ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/PARCC/2015%20PARCC%20Rhode%20Island%20Teaching%20and%20Learning%20Results%20Report.pdf>
- Robertson, P. (2015, June 1). *Peggy Robertson: My comment on PBS News Hour's opt out coverage*. Retrieved from <http://garnpress.com/2015/peggy-robertson-my-comment-on-pbs-newshours-opt-out-coverage/>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. <http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rogers, J., & Mirra, N. (2014). *It's about time: Learning time and educational opportunity in California high schools*. Los Angeles, CA: UCLA IDEA.
- Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39, 406–412. <http://dx.doi.org/10.3102/0013189X10374770>
- Satullo, S. K. (2015, February 20). Pennsylvania part of growing standardized testing opt-out movement. *Lehigh Valley Live*. Retrieved from http://www.lehighvalleylive.com/breaking-news/index.ssf/2015/02/pennsylvania_part_of_growing_n.html
- Severns, M. (2015, July 20). GOP candidates join testing opt-out movement. *Politico*. Retrieved from <http://www.politico.com/story/2015/07/gop-candidates-join-testing-opt-out-movement-120318.html>
- Shanklin, M. (2004, July 30). 18 Orange teachers face forced transfer. *Orlando Sentinel*. Retrieved from http://articles.orlandosentinel.com/2004-07-30/news/0407300341_1_teachers-school-starts-high-school
- Shavelson, R. J., Linn, R. L., Baker, E. L., Ladd, H. F., Darling-Hammond, L., Shepard, L. A., ... Rothstein, R. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 417–453. <http://dx.doi.org/10.3102/00346543075003417>

- Strauss, V. (2013, November 21). NY school principals write letter of concern about Common Core tests. *Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/11/21/n-y-school-principals-write-letter-of-concern-about-common-core-tests/>
- Taylor, K. (2015, March 22). Cuomo fights rating system in which few teachers are bad. *New York Times*. Retrieved from <http://www.nytimes.com/2015/03/23/nyregion/cuomo-fights-rating-system-in-which-few-teachers-are-bad.html?ref=topics&r=0>
- Tens of thousands skip state tests in Illinois. (2016, January 5). *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2016/01/06/tens-of-thousands-skip-state-tests-in.html?qs=illinois>
- Toeh, M., Coggins, C., Guan, C., & Hiler, T. (2014, Winter). *The student and the stopwatch: How much time do American students spend on teaching?* Retrieved from the Teach+Plus website: http://www.teachplus.org/sites/default/files/publication/pdf/the_student_and_the_stopwatch.pdf
- The Editorial Board. (2015, August, 14). Opting out of standardized tests isn't the answer. *New York Times*. Retrieved from <http://www.nytimes.com/2015/08/15/opinion/opting-out-of-standardized-tests-isnt-the-answer.html>
- Tompson, T., Benz, J., & Agiesta, J. (2013). *Parents' attitudes on the quality of education in the United States*. Retrieved from the Associated Press-NORC Center for Public Affairs Research website: http://www.apnorc.org/PDFs/Parent%20Attitudes/AP_NORC_Parents%20Attitudes%20on%20the%20Quality%20of%20Education%20in%20the%20US_FINAL_2.pdf
- U.S. Department of Education. (2009). *Race to the Top Program executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education. (2013a). *The nation's report card mathematics 2013 state snapshot report: New York grade 8 public schools*. Retrieved from <https://nces.ed.gov/nationsreportcard/subject/publications/stt2013/pdf/2014465NY8.pdf>
- U.S. Department of Education. (2013b). *The nation's report card mathematics 2013 state snapshot report: New York grade 4 public schools*. Retrieved from <https://nces.ed.gov/nationsreportcard/subject/publications/stt2013/pdf/2014465NY4.pdf>
- U.S. Department of Education. (2013c). *The nation's report card reading 2013 state snapshot report: New York grade 4 public schools*. Retrieved from <http://nces.ed.gov/nationsreportcard/subject/publications/stt2013/pdf/2014464NY4.pdf>
- U.S. Department of Education. (2013d). *The nation's report card reading 2013 state snapshot report: New York grade 8 public schools*. Retrieved from <https://nces.ed.gov/nationsreportcard/subject/publications/stt2013/pdf/2014464NY8.pdf>
- U.S. Department of Education. (2015, October 24). *Fact sheet: Testing action plan*. Retrieved from <http://www.ed.gov/news/press-releases/fact-sheet-testing-action-plan>
- Ujifusa, A. (2015a, July 8). Amid cries of overtesting, a crazy quilt of state responses. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2015/07/08/amid-cries-of-overtesting-a-crazy-quilt.html>
- Ujifusa, A. (2015b, December 23). Education Department asks 13 states to address low test-participation rates. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2015/12/twelve_states_asked_to_address.html?cmp=eml-enl-eu-news
- Ujifusa, A. (2015c). New York's student opt-out rate for testing hits 20 percent. *Education Week*, 35(1), 4.
- VanLehn, K., & van der Sande, B. (2009). Acquiring conceptual expertise from modeling: The case of elementary physics. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 356–378). Cambridge, England: Cambridge University Press.
- Vernez, G., Naftel, S., Ross, K., Le Floch, K. C., Beighley, C., & Gill, B. (2009). *State and local implementation of the No Child Left Behind Act (Volume VII)— Title I school choice and supplemental educational services: Final report*. Retrieved from <https://www2.ed.gov/rschstat/eval/choice/nclb-choice-ses-final/choice-ses-final.pdf>

Suggested citation:

Bennett, R. E. (2016). *Opt out: An examination of issues* (Research Report No. RR-16-13). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/ets2.12101>

Action Editors: Heather Buzick and James Carlson

Reviewers: Thomas Van Essen and Elizabeth Stone

ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING is a trademark of ETS. All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>