

# FairTest

---

National Center for Fair & Open Testing

## Testimony to House Education Committee Hearing on Accountability - May 7, 2013

**Monty Neill, Ed.D., Executive Director, FairTest**

Mr. Chairman, Members of the Committee,

Thank you very much for allowing this testimony on the vitally important questions of assessment and accountability.

My name is Monty Neill, and I am the executive Director of FairTest, the National Center for Fair & Open Testing. (See <http://www.fairtest.org>.)

Educationally beneficial accountability must rest on strong evidence of important learning outcomes and other strong information about school quality. Accountability must be structured first and foremost to assist school improvement. Where schools clearly are unable to improve, then a healthy accountability system requires stronger actions.

In this testimony I seek to accomplish the following:

- 1) To explain the scope and reasons for a growing parent, student, teacher and public backlash against high-stakes standardized testing, the central component of current accountability systems.
- 2) To summarize briefly and provide references on how and why testing has failed as the key component of accountability.
- 3) To describe and provide examples of what assessment could be, describing systems Congress should help states develop.
- 4) To briefly discuss and provide references on the misuse of testing in the evaluation of teachers. And,
- 5) To outline a step by step accountability structure that Congress could implement to replace No Child Left Behind and the waiver system the administration has persuaded many states to implement.

Please note that I also chair the Forum on Educational Accountability. I am not presenting on their behalf, but I am using a good deal of their work in presenting a superior accountability system. (See <http://www.edaccountability.org>.)

**P.O. Box 300204, Jamaica Plain, MA 02108**  
**fairtest@fairtest.org 617-477-9792 <http://fairtest.org>**

## **Resistance**

We are in a time of rapidly growing resistance to high-stakes misuses of standardized tests. Let me give you some examples:

Providence Students held a zombie walk; they persuaded a group of prominent citizens to take the state test – and most of them failed it. They also released a program for genuine school reform, including use of authentic performance assessments.

In New York City parents and students boycotted testing in 30 schools, then held a 500-person rally. Such boycotts occurred across New York State, where 1600 principals have signed a petition against test misuse, while hundreds of researchers signed an open letter against high stakes testing. Researchers in Massachusetts, Chicago and Georgia have signed similar letters.

In Chicago, students boycotted, parents held a play-in at school headquarters, and a steady flow of public forums on testing have been held across the city. Numerous grassroots community groups have joined with parent and student groups and the Chicago Teachers Union to forge a growing movement. The union made the use of student test scores to judge teachers a key issue in their successful strike last spring, a strike that polls said had the support of a strong majority of the city's people. In response, Chicago has dropped some testing.

In Indiana, voters elected Glenda Ritz, a critic of high-stakes testing, over Tony Bennett, a staunch defender of such testing.

In Seattle, teachers have twice boycotted tests that are not connected to the curriculum and eat up computer labs for a third of the school year, denying students time to use computers for real educational pursuits. Students themselves boycotted the tests when administrators tried to administer them. The administration has reduced the amount of MAP testing – the target of the protests – and educators promise to continue their efforts to ensure reasonable testing.

And in other states, parents and students have ‘opted out’ of the tests, including in California, Colorado, Oregon, Oklahoma, Florida and Pennsylvania.

All this comes on top of the National Resolution on High Stakes Testing, sponsored by FairTest and a dozen others, calling for a sweeping overhaul of assessment and accountability. (See <http://timeoutfromtesting.org/nationalresolution/>). That resolution has been endorsed by more than 530 organizations and 17,000 individuals. Before that was the Texas school boards resolution, which said testing is “strangling” education. It has been signed by 86% of Texas school boards. That has laid the ground for what is likely to be a marked retreat in the amount of mandated testing through bills nearing passage in the Texas legislature.

People are rebelling over the amount of tests, the low quality of the tests, and their misuse as high stakes hurdles for students, teachers and schools.

## **The tests**

I've not the space to discuss in detail the limits and flaws of the tests or the damage caused by their high-stakes misuse. I'll just make a couple of key points, and direct you to references.

First, the tests are narrow and measure only a limited slice of what students need to know and be able to do. High stakes pressures too many schools to teach to the test, narrowing the curriculum and undermining subject quality. This denies children the high-quality education they need and deserve. It is a likely reason why gains in NAEP have slowed and even halted in both subjects, at grades 4 and 8 and 12, for almost every demographic group. Quite simply, the testing mania is not working (Guisbond, Neill and Schaeffer, 2012).

Second, the looming Common Core tests will be, at best, marginally better, a point also raised by the Gordon Commission report. Unfortunately, these new tests have devoured hundreds of millions of dollars and may dominate schooling in the next few years. They will not solve the problem of assessment quality; the high stakes misuses remain; the negative consequences will also continue. (FairTest, 2012a; Gordon Commission, 2013.)

Third, we still have no serious proof that schools can overcome the effects of poverty and racism on a wide scale. Schools continue to account for some 25% of the variance in student outcomes. We should continue to work to improve schools, and perhaps the impact of schools can increase as schools strengthen. But pretending that schools alone will solve poverty, and will do so via standards and tests, is dangerous. It leads us to blame schools and educators for things they cannot possibly accomplish, provides excuses for continuing to poorly fund schools and related programs such as early childhood programs, and allows us to avoid addressing issues of jobs, income, housing, transportation, and other factors that, more powerfully than schools, create the odds of student success.

None of this means we should not assess students, evaluate teachers and schools, gather information that can be used to improve schools, or require no accountability. It means we have failed to construct an educationally sound and healthy way of meeting those important goals.

### **What should we do instead: Assessment**

Over the years, FairTest and its allies have developed a multi-part proposal for assessment and evaluation. It includes limited use of large-scale tests, a core of information from classroom and school evidence, and use of school quality reviews (Neill, 2010).

***Large-scale tests.*** Many nations with better and more equal education outcomes test only one to three times before high school graduation and largely avoid multiple-choice questions (Darling-Hammond, 2010a; FairTest, 2010). Congress should require statewide tests once each in elementary, middle and high school, in language arts and math. Congress could allow states to sample, as NAEP does. The critical point is that no stakes should be attached to these standardized exams. Rather the results would help inform an overall evaluation. Where serious discrepancies exist between test results and other evidence, that could be the basis for an investigation.

***Local and classroom evidence of learning.*** If you want to find out what kids know and can do, look at their actual work. This is what many other countries do (Darling-Hammond, 2010b). By focusing on the classroom, we can assess important learning standardized tests cannot, such as research projects, oral presentations, essays, using computers to solve real-world problems. Such assessments enable us to evaluate higher order thinking skills and deeper knowledge about student

learning than can standardized tests. (Forum on Educational Accountability, ). Developing and using high-quality assessment improves teaching and learning. The evidence can be summed up and presented annually to the school's community and the state (Neill, 2010; FairTest 2010).

Building the system on local evidence means trusting teachers. Some need to improve their assessment skills, so ensuring teachers can work and learn together is important. This is what high-performing nations have done (Darling-Hammond, 2010, a, b).

To ensure quality, some other countries have systems where samples of student work from each classroom are re-scored by independent raters to verify a teacher's initial score ("moderation"). This has been done well enough to ensure local quality and provide comparability across a state. (FairTest, 2010, provides examples and links.) Schools would explain their results in an annual report.

Here I will turn to two examples, as this is the heart of our position: it is feasible to use classroom and school-based evidence in an evaluation process.

The *New York Performance Standards Consortium* includes 26 high schools, 24 of them in New York City, that use a common use a “practitioner-developed and student-focused performance assessment system” (New York Consortium, 2013). They require graduating students to prove their subject knowledge through performance-based assessment tasks that show oral and written skill, including an analytic literary essay, an applied math project, an original science experiment showing understanding of the scientific method, and a history research paper showing valid use of argument and evidence. The tasks are practitioner-based, student-centered.

The students in the New York City Consortium schools are demographically nearly identical to the city as a whole. Their results, which they attribute most strongly to their assessment system, far exceed New York City averages, in terms of graduation rates, college attendance and persistence in college. Indeed, they exceed the national average for percentage of grads still in college in year 3. Test-based, top down education ‘reform’ in New York City, has not worked; the Consortium has.

The Consortium requires that students and teachers work together to develop the topics for the graduation tasks. Each student may have her own task. They are worked on over weeks, not just one or a few periods as with Common Core tasks. They are judged using a common scoring guide across consortium schools. Students must defend their tasks before a committee, including their teacher and two others, usually from outside the school – as do doctoral candidates. Samples are re-scored to ensur consistency across the schools. The system has been independently reviewed and found to be sound.

The *Learning Record* (n.d.) was developed for use with multi-lingual, multi-cultural populations, to assess progress in reading, writing, speaking and listening. Using a structured format, the teacher regularly observes and describes the student and her work, and collects work samples, to provide multiples sources and types of evidence. This is a very detailed process that takes place throughout the school year. The Record is a well-structured instrument that provides clear guidance on how teachers are to document learning across relevant dimensions, from phonemic awareness to deep comprehension, going well beyond what standardized tests measure. Student progress is summarized in writing and the level of learning is placed numerically on a developmental scale. LRs have been re-scored, using hundreds of records, with high inter-rater agreement, and studies have supported its validity.

I would note that there are similarities between the Learning Record and the Work Sampling System developed by Samuel Meisels, which also provides in-depth classroom-based information, can use moderation, and has been validated for use with younger children, ages 3 – 8 (FairTest, 2010).

These are different kinds of systems, but they are complementary. Certainly project-based learning fits with the Learning Record, and Consortium schools often rely on portfolios.

For producing public reporting, for large-scale purposes, the key point here is that teacher judgments can be verified *if* the structures for gathering the work and the processes of evaluation, the scoring guides and procedures, are sufficiently clear and strong.

Moderation, systems of re-scoring and sometimes score adjusting, has been used successfully in other nations (Darling-Hammond, 2010a).

My point is that large-scale moderation rooted in high quality assessment practices can work. This need not be hugely expensive provided that a) moderating becomes a normal part of teachers' work, and b) moderation uses samples. The idea here is that if 3-5 randomly selected Records or Portfolios or common tasks from a given classroom are re-scored, and the teacher's score found accurate, then we can have good confidence in her general accuracy. A placement of '3' on a developmental scale on the Record, or a passing or exemplary score on a task, would mean the same across a wider area, such as a state.

Congress should take steps to dramatically shift course to this direction. However, if Congress is not ready to take such a step, it should authorize a pilot project for states to voluntarily begin constructing truly new, educationally sound assessment systems. I attach a draft amendment toward that end.

Two additional quick points:

First, building this system requires significant professional learning. That is a good thing, because the result is superior teachers, teachers who know their subject, their craft and their students better.

Second, this cannot work in a context of punitive accountability. Evidence, outcomes, must be understood by educators and communities as being used to help improve teaching and schools. Teachers who cannot do their jobs well should be counseled out and if necessary removed, which good teachers will support, and the system I have described will provide far better evidence for that process than do standardized test scores. Similarly, schools that cannot improve despite assistance do require interventions, perhaps including staff removal. But if this is seen as the purpose of the system, or as a quite possible outcome based on bad and erratic data, as it now is in many cities, the system will not work properly.

### ***A note on teacher evaluation***

The preponderance of research evidence shows that tools such as value added and growth formulas cannot be used fairly in judging teachers in real-world contexts (FairTest, 2012b). Efforts to take account of factor such as varying student backgrounds, are inadequate. The very process of using student scores even as a weighted fraction of decisions will be damaging to the life of schools, in part because it will intensify teaching to inadequate tests. The administration should never have required this of states to obtain NCLB waivers, nor should Congress require it when it reauthorizes ESEA (FEA, 2011b).

But, teachers should be evaluated, and student learning should be part of that evaluation. Thus, use of rich forms of evidence of student learning should be included with well-designed reviews and systems such as Peer Assistance and Review. Montgomery County used this well. One consequence was that a significant portion of teachers did leave. Another was that teachers knew they had a tool that was helpful. I use the past tense because Montgomery County believes that imposition of the use of student scores to judge teachers will sabotage what has been a productive system.

### **What should we do: Accountability**

The Forum on Educational Accountability has proposed a fundamentally different approach to accountability. Its work rests on a key point in the Joint Organizational Statement on No Child Left Behind (2004), now endorsed by more than 150 national education, civil rights, religious, disability, parent and civic organizations: *Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.* The Forum (n.d.) has focused on assessment, accountability, school improvement, and equity/opportunity to learn.

FEA's (2011a) recommendations on accountability for the reauthorization of ESEA say:

- Eliminate “adequate yearly progress” (AYP) requirements and sanctions, but continue reporting important data disaggregated by demographic group. Avoid tying goal of ensuring all students are on track to be college and workforce ready to any arbitrary deadline. Expect demonstration of reasonably attainable rates of improvement (e.g., those now achieved by schools in the top quarter on improvement rates).
- In evaluating and recommending interventions in and changes to schools or districts, use both multiple sources of evidence (comprehensive indicators) and periodic reviews of schools and districts by qualified state teams.
- Allow a broad, flexible range of “turnaround” options. Use indicators and reviews to tailor change actions to schools’ needs. Build improvement plans from elements demonstrated to be essential to school improvement--e.g., collaborative professional development, strong leadership, parent involvement, and rich and challenging curriculum--and allow schools and districts to determine how they will address these areas to help build their capacity for long-term improvement.
- Establish the principle of holding schools and districts accountable through monitoring and appropriate public reporting to ensure consistent, successful efforts to fulfill improvement plans.
- Set the percentage of schools required to engage in turnaround activities based on standards for intervention and federal appropriation levels, rather than set percentages regardless of funding.
- Assist states and districts in developing and implementing sound and fair schoolwide evaluation policies aimed at schoolwide improvement, rather than the Blueprint model, which largely shifts test-based accountability from schools to educators. Educator evaluation programs should include evidence of student learning and other measures of educator competency, but the federal government should not mandate the inclusion of scores from large-scale tests.

Further discussion of a few of these recommendations:

FairTest also recommends consideration of school quality reviews (SQR) as a means to accomplish the ongoing school evaluations that FEA recommends regarding accountability and improvement. The SQR is the central tool for school evaluation in places such as England and New Zealand (see Rothstein, 2008; Ratner & Neill, 2009). Instead of test results, their systems

focus on a comprehensive school review by a team of qualified professionals every 4-5 years. This leads to a report describing the school and recommending actions for improvement. Schools that need extra help would be reviewed more frequently. Schools that are reviewed would also provide extensive data about their resources, their processes, how they strive to improve, problems they are encountering, and so on.

This is usually envisioned as a formal process and would be controlled by the state. It may be that states would prefer a more informal process. For example, in England a network, Raising Achievement, Transforming Learning (RATL) pairs schools so they can help each other (Hargreaves and Shirley, 2009). These have been shown to produce improvement. Interestingly, it is not necessary to pair a good school with a weak school – even two weak schools collaborating seem to produce significant improvement. It seems to be the process of thinking and working together that spurs positive changes.

For a more formal evaluation system, the SQR can make a useful contribution by providing rich information beyond evidence of student learning. Schools are not only places of learning, they should be places where children are healthy and happy as well as challenged and supported to learn, in social, emotional and behavioral ways, as well as academically. SQRs, complemented by other sources of information, can provide information for evaluation and more importantly for improvement.

FEA also has developed a proposal for turnarounds based on “common elements” identified by research as key to successful schools and turnaround efforts. FEA recommends that this approach replace the four requirements in Race to the Top and the Department’s NCLB waivers. At a minimum, should Congress retain those four options, then “common elements” should be an additional option. Unlike the Administration’s approach, “common elements” are based on research and evidence from practice. (See Forum on Educational Accountability, 2010; Ratner & Neill, 2010.)

### **In conclusion**

In a period of strong and growing resistance to tests, tests that are educationally inadequate and whose use is failing to genuinely improve America’s schools, as will be the case with the coming Common Core tests, it is imperative that Congress take steps to dismantle the educationally harmful test and punish system it has created.

But we do need to evaluate students, teachers, schools and systems. Schools and districts do need to give an accounting to their communities, the public and the state. The state will need at times to intervene to ensure local officials do their jobs well and schools do their best. And it is fundamentally important to provide educationally sound assistance to schools in need.

The procedures I have described can do that, and do it in ways that are educationally beneficial. FairTest – and FEA – propose a fundamental overhaul of federal law. This now seems beyond

what Congress proposes to do. But we should dream big, our children deserve it. The current system does not work, nor will tinkering solves its dangerous inadequacies. Instead, Congress needs to move in a dramatically different direction. I hope this testimony will help you consider whether and how to do so.

## References

- Darling-Hammond, L. 2010a. *Benchmarking Learning Systems: Student Performance Assessment in International Context*.  
[http://edpolicy.stanford.edu/pages/pubs/pub\\_docs/assessment/scope\\_pa\\_ldh.pdf](http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_ldh.pdf).
- Darling-Hammond, L. 2010. *The Flat World and Education*. New York: Teachers College Press.
- FairTest. 2012a. *Common Core Assessments: More Tests, But Not Much Better*.  
<http://www.fairtest.org/common-core-assessments-more-tests-not-much-better>. Contains further research links.
- FairTest. 2012b. *Why Teacher Evaluation Shouldn't Rest on Student Test Scores*.  
<http://www.fairtest.org/sites/default/files/Teacher-Evaluation-FactSheet-10-12.pdf>. Contains extensive bibliography on this topic.
- FairTest. 2010. *Multiple Measures: A Definition and Examples from the U.S. and Other Nations*.  
<http://www.fairtest.org/sites/default/files/MultipleMeasures.pdf>.
- Forum on Educational Accountability (FEA). n.d. Reports, position papers and recommendations. <http://www.edaccountability.org/>.
- Forum on Educational Accountability (FEA). 2011a. *Recommendations for Improving ESEA/NCLB- Summary*.  
[http://www.edaccountability.org/FEA\\_2\\_Page\\_Summary\\_Recommendations\\_2011\\_final.pdf](http://www.edaccountability.org/FEA_2_Page_Summary_Recommendations_2011_final.pdf).
- Forum on Educational Accountability (FEA). 2011b. "What Should Congress Do About Teacher Evaluation?"  
[http://edaccountability.org/What\\_Should\\_Congress\\_Do\\_About\\_Teacher\\_Evaluation-\\_FEA\\_letter\\_9-20-11.pdf](http://edaccountability.org/What_Should_Congress_Do_About_Teacher_Evaluation-_FEA_letter_9-20-11.pdf).
- Forum on Educational Accountability (FEA). 2010. *A Research- and Experience-Based Turnaround Process*. <http://www.edaccountability.org/pdf/FEA-TurnaroundStatementJune2010.pdf>.
- Forum on Educational Accountability, Expert Panel on Assessment. 2007. *Assessment and Accountability for Improving Schools and Learning*.  
<http://www.edaccountability.org/AssessmentFullReportJUNE07.pdf>
- Gordon Commission on the Future of Assessment in Education. 2013. *A Public Policy Statement*.  
[http://www.gordoncommission.org/rsc/pdfs/gordon\\_commission\\_public\\_policy\\_report.pdf](http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf).

Guisbond, L., Neill, M., & Schaeffer, B. 2012. *NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure?*

[http://fairtest.org/sites/default/files/NCLB\\_Report\\_Final\\_Layout.pdf](http://fairtest.org/sites/default/files/NCLB_Report_Final_Layout.pdf). Contains relevant, extensive bibliography.

Hargreaves, A., and Shirley, D. *The Fourth Way*. Thousand Oaks, CA: Corwin.

Joint Organizational Statement on No Child Left Behind. 2004.

[http://www.edaccountability.org/Joint\\_Statement.html](http://www.edaccountability.org/Joint_Statement.html).

Learning Record. N.d. <http://learningrecord.org>.

Neill, M. 2010. "A Better Way to Assess Students and Evaluate Schools," *Education Week*, June 18. <http://www.edweek.org/ew/articles/2010/06/18/36neill.h29.html>.

New York Performance Standards Consortium. 2013. *Educating for the 21st Century: Data Report on the New York Performance Standards Consortium*.

[http://performanceassessment.org/articles/DataReport\\_NY\\_PSC.pdf](http://performanceassessment.org/articles/DataReport_NY_PSC.pdf). More about the Consortium can be found at <http://performanceassessment.org/>.

Ratner, G., and Neill, M. 2010. *Common Elements of Successful School Turnarounds: Research and Experience*.

<http://www.edaccountability.org/pdf/CommonElementsSuccessfulSchoolTurnarounds.pdf>.

Documents the research evidence behind FEA 'common elements' recommendations.

Ratner, G., and Neill, M. Integrating 'Helping Schools Improve' With 'Accountability' Under ESEA: "The Key Role For Qualitative, As Well As Quantitative, Evaluations And The Use Of "Inspectorates," Working Paper II. [http://www.fairtest.org/sites/default/files/SQR-Inspectorate\\_working\\_paper\\_2.pdf](http://www.fairtest.org/sites/default/files/SQR-Inspectorate_working_paper_2.pdf).

Rothstein, R. 2008. *Grading Education: Getting Accountability Right*. Washington and New York: Economic Policy Institute/Teachers College Press. See esp., Ch. 7.