

FairTest

National Center for Fair & Open Testing

FairTest here elaborates on its general comments by responding in detail to the draft Guidelines for states to apply for "Race to the Top" (RTTT) funds.

I. Proposed Priorities.

Problem: The proposed priorities have little to do with improved curriculum, instruction, assessment, professional development or school improvement, except under priority 2, which focuses only on STEM. Priorities 3-5 are about data and management. Under priority one, a state could emphasize improving curriculum, instruction, assessment, and professional development, but there is too little incentive in the Requirements or Selection Criteria to do so (as explained below). Thus, the guidelines place little emphasis on what should be central to improving schools and great emphasis on points that are less central, though some could be useful.

Solution: Revise the guidelines to prioritize development and implementation of core school improvement activities, particularly school-based collaborative activities to improve teaching. The Forum on Educational Accountability has described several key points that are central to improving schools' capacities to serve all their children well. (See <http://www.edaccountability.org>.)

II. Requirements.

A. Eligibility Requirements

Problem: The text reads: " Second, we propose that to be eligible under this program, a State must not have any legal, statutory, or regulatory barriers to linking student achievement or student growth data to teachers for the purpose of teacher and principal evaluation... Therefore, one of the most effective ways to accurately assess teacher quality is to measure the growth in achievement of a teacher's students [4,5]; and by aggregating the performance of students across teachers within a school, to assess principal quality... This capability is fundamental to Race to the Top."

The guidance suggests the only choice available to the Department is between relying only on things like background (certification, degrees) or using student test scores. This leaves out school-based evaluation procedures (c.f., the work of Charlotte Daniels, the National Staff Development Council, and Linda Darling-Hammond). The "therefore" in the text does not follow, it is a non sequitur. While the draft makes a case (in footnotes) for not relying only on things like diplomas, it makes no case for requiring a linkage between test scores and teacher evaluations, never mind salary, tenure and firing decisions, as the guidelines propose (below).

(more...)

15 Court Square, Suite 820, Boston, MA 02018 (857) 350-8207 FAX (857) 350-8209
Web Site: www.fairtest.org Email: fairtest@fairtest.org

The American Recovery and Reinvestment Act (ARRA) itself simply says:
"(2) ACHIEVING EQUITY IN TEACHER DISTRIBUTION.—The State will take actions to improve teacher effectiveness and comply with section 1111(b)(8)(C) of the ESEA (20 U.S.C. 6311(b)(8)(C)) in order to address inequities in the distribution of highly qualified teachers between high- and low-poverty schools, and to ensure that low-income and minority children are not taught at higher rates than other children by inexperienced, unqualified, or out-of-field teachers."

The approach mandated in the draft guidelines is not at all fundamental, but rather one possible (though unproven by research) means to address "teacher effectiveness." The linkage also is not necessary for addressing the second part of the law's requirements, "inequities in the distribution of highly-qualified teachers," at least as highly-qualified is defined in ESEA.

The draft guidelines continue: "Without this legal authority, States would not be able to execute reform plans relating to several selection criteria in this notice (see Selection Criteria (C)(2) through (C)(5)), because these plans must require LEAs and schools to determine which teachers and principals are effective using student achievement data."

This is circular reasoning: the criteria the Department has chosen (linking student scores to educator evaluations) are not necessary. They merely build on the non sequitur noted above. If this linkage is not mandated, then there is no need to pressure states to change their laws.

Given there are 19 "Selection Criteria," it seems unlikely a state must meet all of them. Nothing in the selection criteria says states *must* construct systems to use teacher evaluations to determine salaries, grant tenure or fire teachers. Indeed, the guidelines themselves say the Department *may* apply one or more of the criteria... (Selection Criteria, emphasis added). In sum, the Department has mandated the removal of barriers to using student achievement data for evaluating teachers and principals, without providing any evidence that this action will be beneficial or desirable.

(These points also will apply to the responses to relevant Selection Criteria.)

Solution: Remove this requirement entirely from the guidelines.

III. Selection Criteria.

(A)(1) *Developing and adopting common standards:*

Problem: For phase II funds, states must have adopted the common standards that will not be ready until the end of 2009 (at earliest). The application will be submitted by 'late spring.' This gives states little time to adopt them, which may require legislative action, and little time for a state to consider the consequences and desirability of those standards, including whether they are an improvement over a state's existing standards.

Solution: *Do not link state willingness to adopt the new standards to the ability to receive ARRA funds. Not linking funds would make state participation truly voluntary.*

(A)(2) *Developing and implementing common, high-quality assessments:*

Problem: National or common test proponents argue that "internationally benchmarked" or "tougher tests" will improve student learning. But tougher tests do not necessarily lead to

better education. While South Carolina's tests are as difficult as Massachusetts', South Carolina has not shown much progress on the National Assessment of Educational Progress (NAEP), though Massachusetts has. Montana, Idaho and Wyoming have very similar NAEP scores and gains, but the difficulty of their tests varies considerably. Something other than "tough tests" must explain these results.

Internationally, top-ranked nations such as Finland, Hong Kong and Singapore do far less testing with far lower stakes. Finland and Hong Kong use occasional sampling, such as NAEP. Singapore relies on classroom-based evidence, except for a grade six test. None mandates standardized testing across multiple grades and none has high-stakes tests for students, schools or teachers in earlier grades. These nations do well by focusing on the quality of the teaching force and the curriculum.

ARRA itself does not require consortia or "international benchmarking." It does call for improved assessments.

Solution: Drop this requirement. Address the need to improve assessments within what is now (A)(3).

(A)(3) Supporting transition to enhanced standards and high-quality assessments:

Discussion: College and career readiness are reasonable goals, provided that career readiness is not reduced to college-oriented content and skills and that areas such as technical-vocational education are assessed under their own standards or criteria. The definition of "high quality assessment" in the guidelines is reasonable, those for "formative" and "interim" assessments are adequate. The list of examples for state options includes things that are fine, but the list is too limited.

Problem: The guidelines do not explicitly encourage states to develop more comprehensive assessment systems that would utilize local and classroom-based evidence of student learning for public reporting, accountability and improvement.

Solution: The guidelines should explicitly allow states to develop assessment systems that include substantial local assessments for use public reporting, accountability and improvement, per the recommendations of the Forum on Educational Accountability (FEA) and as found in the draft legislation for ESEA reauthorization offered by Reps. George Miller and Buck McKeon in 2007.

- The Miller-McKeon draft said:
“SEC. 1125. PILOT PROGRAM TO INCLUDE LOCALLY DEVELOPED MEASURES.
“(a) PILOT PROGRAM ESTABLISHED.—The Secretary may carry out a pilot program under this section 4 under which up to 15 States may include, as part of the assessment system and in addition to State assessments described in section 1111, locally developed, classroom-embedded assessments. Such assessments may be different across local educational agencies and such assessment systems may be used for the purposes of determining adequate yearly progress under section 1111(b)(2).”

The section details the provisions for the pilot program. This language can and should be used by the Department to encourage innovative, comprehensive assessment reforms.

- The *Joint Statement on No Child Left Behind* (signed by 151 national organizations) says: "Help states develop assessment systems that include district and school-based measures in order to provide better, more timely information about student learning."

- FEA's *Empowering Schools and Improving Learning* (signed by 84 national organizations) elaborates: "5. *Assessment*: ESEA shall provide funds to enable schools, districts, and states to develop high quality formative and summative assessments in the various subjects, as well as other indicators to provide evidence of improved student learning and school quality. These assessments must be based on state standards and the local curriculum, assess higher order thinking and other 21st century skills, and provide multiple approaches for students to demonstrate their learning. The primary use of these assessments is to improve instruction and enable teachers to better address each student's strengths and needs. These funds may be used jointly with the funds authorized for collaborative activities and professional development in the school and school district, provided those activities include developing assessments and indicators and improving educators' skills in using them. Since the federal government has previously provided substantial funds for the improvement of statewide standardized tests, Congress should continue modest funding for those activities, which also may include the development of tasks and projects (performance tasks) that states, districts, and schools can use. In addition, Congress shall provide funds for the use of universal design principles to create large-scale and classroom-based assessments that are appropriate for all students, including English language learners and students with disabilities."

- The *Joint Statement, Empowering Schools* and supplemental materials are on the web at <http://www.fairtest.org> and <http://www.edaccountability.org>.

(C)(2) Differentiating teacher and principal effectiveness based on performance:

Problem: The comments at II.A. Eligibility Requirements pertain here. There is no evidence to show that schools or student learning will improve if the linkage between "student achievement" and evaluation, payment, tenure of firing of teachers and principals is established. International research shows it leads to narrowing the curriculum and teaching to the test, to the detriment of all-around learning.

"Achievement" is defined for subjects for which NCLB requires testing as "at a minimum" test scores, and for other subjects test scores are prioritized. The proposal also calls for using "student growth" (defined – at state option -- as entirely or mainly test scores) as a "significant factor."

Test scores, however, provide very inadequate information about student learning. There are too few questions on any topic, the format is too narrow, and they shed little light on what is not working well or how to get better. State exams only weakly assess what is in state standards, usually the lower levels of content and thinking, so that many important areas of learning are not measured.

By insisting on linking educator evaluations to this definition of "achievement," the guidelines intensify the focus on raising those scores. However, there is strong evidence that the greater the stakes – and these are very high stakes -- the more likely are teaching to the test and narrowing the curriculum. These undermine the quality of instruction and the curriculum,

damaging the educational opportunities of many children. Thus, the perverse consequence of tying educator evaluations to test scores will be to undermine teaching, learning and school improvement.

The guidelines themselves recognize this problem by calling for improved assessment. It is not necessary to wait for perfect assessments, but the quality of current assessments is so low, they should not be used as high-stakes tools in evaluating educators.

In addition, researchers have concluded that similar sorts of problems in other professional fields, such as medicine, which helps explain why payment for results is quite rare. Among those who have documented the negative consequences and rare use of payment for results are Adams, Heywood and Rothstein, *Teachers, Performance Pay, and Accountability*; and Madaus, Russell and Higgins, *The Paradoxes of High Stakes Testing*.

Solution: The criterion of linking achievement to evaluation should be removed. If that cannot be removed, then it should be clarified that anything beyond "evaluation" is not at all necessary, and that no penalty will be exacted against or points awarded to states that do not propose to use test scores to determine payment, tenure or dismissal.

It is appropriate for the guidelines to call for states to submit proposals to develop comprehensive, high-quality evaluation systems. Those systems would look at the full range of desired learning outcomes, with test scores only one small part, as well as the many other important things teachers and principals do.

C)(4) Reporting the effectiveness of teacher and principal preparation programs:

Problem: Parallel to the problems with (C)(2), this criterion will pressure colleges of education to train teachers in more effective test preparation and encourage them to narrow their curriculum. As a consequence, a range of potentially effective pedagogies for broader and richer learning will be reduced or dropped.

Solution: Support states in building data systems that incorporate the multiple facets of high quality education, then subsequently (not under these guidelines) use those systems as part of a process of evaluating teacher training programs.

(C)(5) Providing effective support to teachers and principals:

Problem: Providing effective support is essential to school improvement. However, this criterion focuses on "rapid-time" responses (responses made within 72 hours) rather than on building comprehensive professional learning systems and integrating those into building the capacity of all schools to serve children well. "Rapid-time" (as defined in the guidelines) is far too limited a concept and practice to support the longer-term improvement tasks that are identified in this criterion. The language blurs "rapid-time" with a range of potentially useful activities, suggesting a reduction of professional learning to "rapid-time."

The use of formative assessments can be part of rapid-time responses, and would be a useful component of school improvement. However, the key professional development and school improvement issues are how educators can craft and effectively use such assessments.

Real improvement will have to be a largely local activity, supported by states. The language here can foster the illusion that states themselves can do the detailed school improvement work or even do high-quality "rapid-time" efforts when such efforts are a reasonable component of improvement work.

Overall, there is far too little in the guidance that will actually help improve schools.

Solution: Recast this criterion to focus on support for comprehensive professional learning and supports for teachers and principals. Understand this must be primarily a local effort, but with state support. States can reasonably expect local districts to explain what they intend to do and to then document what they did, how well it worked, and how they would improve what they do. The Forum on Educational Accountability's *Redefining Accountability* provides an outline of how the federal government can approach this issue (available at <http://www.edaccountability.org>). Remove "rapidtime" as an explicit component of professional learning efforts; states and districts can include any such approach if they see it as making sense.

(D)(3) Turning around struggling schools:

Problem: While blocking the more flexible "other major restructuring" option allowed by NCLB that some states are using successfully, it continues the law's automatic requirement to take extreme, often ineffective actions based solely on test scores.

There is little to support the approaches this criterion will mandate. Charters overall are not better than regular public schools, and EMOs have not proven effective. Sometimes replacing principals is a good idea. Closing a school and placing students in "high performing schools" (another specified option) will rarely be possible. Direct state takeovers, which is included in No Child Left Behind, is not on the list. in these bullets. The last bullet in this criterion lists actions to take on a school when the other "strategies are not possible." No criteria are given for defining "not possible." This section appears to replace NCLB's "any other major restructuring" – but it is far less flexible and is only allowed when the other specified options are "not possible."

Evidence from the Center on Educational Policy's research on how states are implementing NCLB provides no justification for narrowing the options. Indeed, schools that are doing poorly and not improving should be the focus on major attention, and specifying the bottom 5 or 5 percent is reasonable. However, the improvement proposals here lack evidence they will succeed and reduce state flexibility.

Solution: Revise this section to call on states to more clearly specify how they will meet NCLB requirements. Maintain the 5 or 5 percent requirement for more focused actions. The Department may conclude it is beyond its authority to end NCLB's test-based automatic requirements for governance-related actions, but it should avoid reinforcing the use of a limited set of unproven options based almost entirely on standardized test scores.

IV. Definitions.

The following definitions must be modified in line with the comments made above, particularly in regard to the overemphasis on standardized tests that shape many of these definitions:

- Effective principal;
- Effective teacher;
- Highly effective principal;
- Highly effective teacher;
- Rapid-time;
- Student achievement; and
- Student growth.