

Portfolios: A Backward Step in School Accountability

By Robert Holland

September 2007



1600 Wilson Boulevard, Suite 900
Arlington, VA 22209

Tel : 703.522.5828 Fax : 703.522.5837

www.lexingtoninstitute.org mail@lexingtoninstitute.org

Portfolios: A Backward Step in School Accountability

By Robert Holland

September 2007

Executive Summary

With Congress set to take up the thorny issue of reauthorizing the No Child Left Behind Act (NCLB), lobbyists for the National Education Association and several other education interests are advocating a major change in accountability for the billions spent on NCLB. Instead of states being required to administer annually a standardized test of students' ability to read and to compute, they want states to apply "multiple measures," including, most notably, portfolio assessment.

Portfolios are collections of student work, such as essays, artwork, and research papers. Progressive educators long have advocated that portfolios be substituted for paper-and-pencil tests because they are more "natural" and "authentic."

In the 1990s, Vermont and Kentucky implemented portfolio assessment as an integral part of education reform plans. Separate studies by nationally respected researchers showed that as a school accountability tool, portfolio assessment was a huge flop in both states, yielding results that were wildly unreliable and very expensive to obtain.

Among the problems found:

- A failure to yield reliable comparative data.
- Large differences in the way teachers implemented portfolios.
- Major differences in the degree of difficulty of assignments, rendering comparisons among students or groups of students highly misleading.

The question is why anyone sincerely interested in holding schools accountable for results would want to revive such a failed method of assessment. Details follow.

Portfolios: A Backward Step in School Accountability

By Robert Holland

September 2007

The following exchange occurred July 30, 2007 at the National Press Club following an address by House Education Committee Chairman George Miller, Chairman of the House Education Committee, on the pending reauthorization of the federal No Child Left Behind Act (NCLB):

QUESTION: You had talked a moment ago about using portfolios possibly for students with limited English. Would those be portfolios solely for . . .

MILLER: I just walked off into deep water. No, not solely for it, no, no, no, no.

(LAUGHTER)

QUESTION: But . . . students sort of --- would this be . . .

MILLER: There's been a discussion in this country for 50 years about portfolios. I won't get into my characterization. . .

QUESTION: So where would you . . .

MILLER: I'll have to get back to you on the details of that. I don't want to – it's a very delicate subject

It was fitting that the subject of portfolio assessment as a possible measure of student progress under a retooled NCLB caused such a stir at the venerable Press Club and prompted Chairman Miller to backpedal and exercise an abundance of caution.

Portfolio assessment does have a history (actually, more like 100 years than 50) and it has a track record as a method of statewide evaluation of student achievement. That record yields no grounds for believing portfolios would be a valid method of gauging whether schools have done their job helping children master basic skills that are critical to their academic success.

The NEA's Push for Portfolios

Yet, the National Education Association (NEA), the nation's largest teacher union, is applying pressure on Capitol Hill to revise NCLB so that portfolio assessment would

become one of the preferred ways for states to prove that they are getting results from the billions of tax dollars being funneled to them under NCLB.

In March 2007, the NEA released its list of “Top Legislative Priorities” for the reauthorization of No Child Left Behind, the new and improved version of the 1965 Elementary and Secondary Education Act that a strong bipartisan majority of Congress with the support of President George W. Bush enacted in 2001. At the top of the NEA’s wish list for a revised NCLB annual accountability system was “inclusion of multiple measures . . . so that AYP (adequate yearly progress) is not based solely on standardized test scores.”

This is what the NEA proposed specifically: “Each state shall incorporate into its accountability system additional measures beyond the current use of just two statewide test scores. Other measures states could include are: district-level assessments, graduation rates for high school, attendance rates, school-level assessments, **performance or portfolio assessments** (emphasis ours), and the percent of students participating in rigorous coursework, which may include dual enrollment, honors, AP (Advanced Placement), or IB (International Baccalaureate) courses.”¹

In a statement to a congressional committee in March, NEA President Reg Weaver said the union wanted state assessments to “include more than what can be assessed on a paper-and-pencil multiple-choice test,” because scores on such tests “reflect little more than a student’s ability to regurgitate facts.”²

Weaver’s remarks were consistent with a steady drumbeat of NEA criticism of objective testing down through the years. Given the NEA’s opposition to such reforms as value-added assessment linked to merit pay for teachers, it is reasonable to suspect that the 3.2 million-member union wants more subjective forms of testing in order to conceal the reality that some schools are failing to teach children vital basic skills of literacy and computation. Under the existing version of No Child Left Behind, states must test students annually in grades 3-8 and once in high school for their mastery of basic reading and math skills.

Concessions to the NEA?

In his prepared address to the Press Club, Chairman Miller, who has been a strong supporter of NCLB accountability, seemed to be yielding ground to the NEA. He said assessments should “include multiple measures of success.” He said graduation rates had to be one such measure. And he implied that some form of performance or portfolio assessment might have to be part of the equation, too. For instance, he said English language learners and disabled students might be among the beneficiaries of new kinds of tests.

Tests must “no longer reflect just basic skills and memorization, but rather critical thinking and the ability to apply knowledge to new and challenging contexts,” Chairman Miller continued.

“To achieve this goal, schools must no longer prepare our students to be autonomous problem solvers. The workplace they enter tomorrow will increasingly require them to work in teams, collaborating across companies, communities, and continents. They will have to draw on different sources of information to solve problems and assemble solutions. These skills cannot be developed solely by a simple multiple-choice exam,” said Miller.³

Well, of course they can't. However, the much-maligned standardized test is an efficient way to find out if a child has acquired the basic tools of literacy that will enable him or her to go on to the higher thinking and collaboration. Before joining a team, it is necessary to become a competent individual. Before solving complex problems, it is necessary to be able to read, write, and compute.

Workforce Imperative?

The NEA is not the only powerful organization pushing for a shift from basic-skills testing to performance-based or portfolio measures. On December 14, 2006, Marc Tucker's New Commission on the Skills of the American Workforce released “Tough Choices or Tough Times,” a warmed-over version of the original Commission's 1990 report seeking to reinvent public education to make it a tool of global workforce development. That report was among those used to push a new paradigm for education called Outcome-Based Education.

In his latest report, Tucker, the president of the National Center on Education and the Economy, writes that nothing less than a “major overhaul of the American testing industry” is necessary. Such a radical revamping, he said, should ensure that tests no longer primarily measure “the acquisition of discipline-based knowledge in the core subjects of the curriculum,” but rather gauge such qualities as “creativity and innovation,

There is no evidence that portfolios provide valid data for measuring the overall performance of a school or a school district.

facility with the use of ideas and abstractions, the self-discipline and organization needed to manage one's work and drive it through to a successful conclusion, the ability to function well as a member of a team, and so on.”⁴

The problem with that assertion, as education scholar Diane Ravitch pointed out in *Education Week*, is that the Tucker report “fails to identify a single test, not in this country or anywhere else, that successfully measures these qualities.”⁵

The Portfolio Record

What about portfolio assessments, the way of measuring that Tucker, the NEA, and assorted progressives (for example, a group calling itself the Forum on Educational Accountability) tout as “authentic”? This approach entails evaluating students according

to collections of their work (such as essays, solved problems, lab reports, and research projects) they have been required to keep as evidence of their talents.

Unfortunately for their advocates, there is no evidence that portfolios provide valid data for measuring the overall performance of a

“The portfolio program is highly unstandardized in almost all aspects of its operation.”

school or a school district. They may be a useful motivational device for a classroom teacher working with individual students, but judging the worth of collected reports about disparate topics is inherently too subjective to be a basis for large-scale assessment.

That conclusion is not a matter of opinion. It is based on many years of experience and solid evidence, as the balance of this report will document.

Roots in Progressive Education

As education writer Jay Mathews documented in a 2004 *Education Next* article, the concept of authentic or portfolio assessment goes back at least 100 years, to the early years of the progressive education movement. Although the method was cumbersome, it fit well with the progressives’ “emphasis on cultivating research skills and creative thinking rather than building a broad base of knowledge in the subject.”

Mathews further noted that portfolios became a part of many alternative public schools that came into being in the 1960s and 1970s. In the 1980s, the Educational Testing Service and scholars from Harvard University collaborated on an experiment using portfolios for students’ writing, music, and art; however, the emphasis was on specific teachers and students, not ways to measure accountability.⁶

With the momentum gathering for standards-based reform in the 1990s led by the nation’s Governors and the first President Bush, Vermont and Kentucky stepped forward as the most enthusiastic proponents of substituting portfolio assessment for standardized testing in evaluating whether schools, school districts, and the states as a whole were making adequate progress.

Authoritative evaluations in the mid-1990s of both states’ ventures in portfolio assessment revealed substantial flaws. Those findings were so compelling that they should give pause to federal or state policymakers before they attempt to dust off portfolio assessment for No Child Left Behind accountability.

Lack of Validity in Vermont

In Vermont, a RAND Corporation team led by respected researcher Daniel Koretz found that portfolio assessment failed to yield reliable comparative data about student achievement. One large problem was the “large variations in teachers’ implementation.” For example, “if two students produce similar work, but one of them has been given greater opportunities for revision or more highly structured help in revising the product,

the similarity of final products is misleading. Likewise, if teachers assign variants of the same problem that differ markedly on difficulty, comparison among their students would be deceptive.”

The RAND study also found that any positive effects had come “at a steep price in time, money, and stress.” It continued:

“The non-financial costs reported by educators – in particular, the time teachers devoted to the program and the sometimes severe tensions the program

The RAND study also found that any positive effects had come “at a steep price in time, money, and stress.”

created – were large, and they represented only the initial stages of a continuing and still difficult developmental effort. The financial costs also appear to be sizable. Indeed, the fact that the Vermont program is still limited to two subjects (writing and mathematics) and two grades is in part testimony to its costliness and difficulty.”⁷

Kentucky’s Flaws

In 1995, the Office of Educational Accountability, Kentucky General Assembly, released an analysis by nationally respected testing experts of that state’s portfolio assessment system. The panel found flaws similar to those uncovered in Vermont.

In 1999, as part of a complex analysis attempting to establish a template for legislatively establishing education standards and assessments, Lance Izumi of the Pacific Research Institute looked closely at the 298-page Kentucky report. This is what he found:

“In 1990, Kentucky passed an education reform act that required portfolio assessment of students, but an academic review panel found huge problems in the system. According to the panel’s findings, large scoring gains by Kentucky students on the portfolio assessment failed to be matched by similar gains on other tests such as the National Assessment of Educational Progress, which strongly suggested that the test results were invalid.

“Also, use of ad-hoc judgmental procedures for linking or equating assessments from year to year likely resulted in making the year-to-year comparison of scores invalid. Further, because teachers in the students’ own schools scored the portfolios, results were biased upward by a significant amount.

“There were also no control factors affecting reliability such as initial instructions from teachers; the amount and type of pre-teaching using similar tasks; time allotted to performance of the tasks; amount and type of assistance from teachers, peers, parents, and others in performing the tasks; opportunities for revision; and the amount and type of assistance in revision provided by others. Because of these problems, Kentucky has reinstated a standardized norm-referenced test.”⁹

Subjectivity

In one section of its report, the reviewers of the Kentucky assessment system, led by Dr. Ronald Hambleton of the University of Massachusetts/Amherst, gave the flavor of just how absurdly subjective the attempted use of portfolios for high-stakes assessment can be:

“The portfolio program is highly unstandardized in almost all aspects of its operation. There are very general guidelines about the types of tasks that should be included; for example, the writing portfolio in grade 8 should include one piece that achieves any one or combination of seven different purposes, such as defending a position, predicting an outcome, or creating a model. The writing projects prohibit the inclusion of group products, while the mathematics guidelines allow one (although both in fact permit and even encourage substantial assistance from other people in performing and revising any portfolio tasks).

“But beyond those general guidelines, teachers and students are largely unrestricted in their selection of tasks and have substantial leeway in terms of the types of products they entail. The mathematics guidelines, for example, note that videotapes, audiotapes, and computer disks are all acceptable products. Tasks may differ among schools or over time in terms of difficulty, novelty, time allowed for completion, amount of revision permitted, and so on.

“Policies pertaining to revision of portfolio pieces are highly permissive. For example, the writing guidelines note that peers may offer suggestions for revisions and call for teachers to instruct students on how to review each others’ writing. They also note that teachers may provide revision check lists. Similarly, the mathematics guidelines also mention parental input but do not specify what that input should be. . . .”¹⁰

As testing expert Richard Phelps noted in his 2005 book, *Kill the Messenger*, not only are portfolios “notoriously difficult to score in a consistent, standardized manner,” but also “they are open invitations to cheating. If a student turns in a portfolio for a statewide, standardized assessment, who is to know if the essay enclosed is written by that student, or her mother, or by someone in New Zealand who posted it on the Internet?”¹¹

Education Fads Die Hard

As Phelps also noted, portfolios are “popular with radical constructivists enamored with the idea of naturalness in education.” Now it is clearer than ever that progressive educators’ infatuation with portfolios (as with other fads) does not die easily.

“Unfortunately, and against all good sense, portfolios remain part of the Kentucky accountability system,” retired University of Louisville professor George K. Cunningham informed us recently. Specifically, Kentucky still uses portfolios as one factor in holding schools accountable for teaching writing.

Dr. Cunningham serves on the School Curriculum and Accountability Council (SCAAC), a panel appointed by the Governor to advise the State Board of Education. Recently, he said, the panel rejected his bid to de-emphasize portfolios in adding up the writing scores. Several years ago, the SCAAC voted to eliminate portfolio scores from accountability, but the head of the program emotionally (and successfully) pleaded to retain them.

At one time, Kentucky also required math portfolios, “but they could not be made to work although the Education Department tried mightily to use them,” Cunningham said. The powerful House Education Committee “finally had to threaten to cut off funding to get them to stop.” Cunningham added that “portfolios have dismal reliability and they take up an inordinate amount of teacher time and resources.”

Having students keep portfolios of their writing can be a useful technique on an individual level. But Cunningham points to major problems with using writing portfolios for statewide accountability. Notably: “Teachers have to play a big part in directing the creation of student portfolios and at the same time they are being evaluated by the portfolios. This leads to the tendency of teachers to be too helpful and for the portfolios to become more of a teacher than a student product. To counter this, Kentucky enforces draconian rules about how

much teachers can help students. In effect, teachers are forbidden to correct student work. . . .”¹²

“. . . portfolios have dismal reliability and they take up an inordinate amount of teacher time and resources.”

That would not seem to be a rational way to improve student writing, although of course some progressives argue against correcting student work, contending it should be entirely about expressing feelings freely.

High-Priced Folly

Finally, a not inconsiderable problem with portfolio assessment is its cost. The RAND study of Vermont’s system noted that there had been no accounting of the financial costs, “but it is clear that they are large – far larger than some optimistic reports have projected.” It continued:

“At the state level, even routine operational costs are large. For example, we estimate that only two of the costs of scoring in 1993 – honoraria and room and board for raters – amounted to about 13 dollars per mathematical portfolio. This \$13 is only a small fraction of state-level operational costs, of course. Scoring entails many other costs (such as substantial staff time both before and during scoring, scanning and analysis of scores, and so on). . . .”¹³

Conclusion

It is difficult to comprehend why any consideration is being given to reviving portfolio assessment as a way to gauge the effectiveness of No Child Left Behind, given the well-documented experiences with this technique of measurement in Vermont and Kentucky during the 1990s. Well-respected researchers found these failings, among others, for portfolios' use in high-stakes assessment:

- A failure to yield reliable comparative data.
- Large differences in the way teachers implemented portfolios.
- Major differences in the opportunities students were given to revise their work, resulting in misleading data when students' collected work was compared.
- Great differences in the degree of difficulty of assignments, rendering comparisons among students or groups of students highly misleading.
- A high price exacted in money, time, and stress on staff.
- A lack of control factors, such as teachers' initial instructions to students.
- Variations in the degree of assistance students amassing their work portfolios receive from peers, parents, teachers, and other sources.
- Opportunities for cheating by importing work not one's own.

Of course, portfolios can be perfectly wonderful devices for classroom teachers to use with individual students. By looking at the collected work (essays, art, research papers, etc.) over a year, teacher and student can reflect on what has been gained and where improvements still can be made. Moreover, teachers can show parents on parent conference day what their children have been doing.

For large-scale evaluation, on the other hand, standardized tests primarily using multiple-choice questions that can be machine-scored offer education officials the best value in terms of reliability, accuracy, ability to generalize the results, ease of scoring, and costs. Obviously, education is about much more than test-taking. But to test once a year to ensure that kids are learning to read and compute up to an acceptable standard does not seem to be an unreasonable requirement.

* * * * *

Endnotes:

¹ NEA's Top Legislative Priorities for ESEA, March 21, 2007; available at: <http://www.nea.org/eSEA/legpriorities.html>

- ² Statement of NEA President Reg Weaver before the Senate and House Education Committees, March 13, 2007; available at:
http://help.senate.gov/Hearings/2007_03_13/Weaver.pdf
- ³ Remarks of House Education Committee Chairman George Miller on the Future of the No Child Left Behind Law, National Press Club, July 30, 2007; available at:
http://www.house.gov/apps/list/speech/edlabor_dem/RelJul30NCLBspeech.html
- ⁴ “Tough Choices or Tough Times,” Report of the New Commission on the Skills of the American Workforce, December 14, 2006; available at:
<http://www.skillscommission.org/report.htm>
- ⁵ Ravitch, Diane, “‘Tough Choices’: Radical Ideas, Misguided Assumptions,” *Education Week*, January 17, 2007; available at:
<http://www.edweek.org/ew/articles/2007/01/17/19ravitch.h26.html>
- ⁶ Mathews, Jay, “Whatever Happened to Portfolio Assessment: Can it be used to hold schools accountable?,” *Education Next*, 2004, No. 3; available at:
<http://www.hoover.org/publications/ednext/3261856.html>
- ⁷ Koretz, Daniel, Stecher, Brian, Klein, Stephen and McCaffrey, Daniel, "The Vermont Portfolio Assessment Program," *Educational Measurement: Issues and Practice*, Fall 1994: 12-13; available at: <http://www.rand.org/pubs/reprints/RP366/>
- ⁸ Hambleton, Ronald K., Jaeger, Richard M., Koretz, Daniel, Linn, Robert L., Millman, Jason, and Phillips, Susan E., "Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991-1994," Office of Educational Accountability, Kentucky General Assembly, 20 June 1995: 4; available at:
<http://www.lrc.ky.gov/oea/reports/MEASUREMENT%20QUALITY%20FINAL%20REPORT%2091-94.pdf>
- ⁹ Izumi, Lance T., “Developing and Implementing Academic Standards,” Pacific Research Institute, May 1, 1999; available at:
http://www.pacificresearch.org/publications/id.240/pub_detail.asp
- ¹⁰ Hambleton, Op. cit.
- ¹¹ Phelps, Richard P., *Kill the Messenger; The War on Standardized Testing*, Transaction Publishers, 2003 and 2005.
- ¹² Cunningham, George K, “Portfolios in Kentucky,” e-mail interview for this report, August 15, 2007.
- ¹³ Koretz, Op. cit.