



High Stakes Exams

by Sarah Warshauer Freedman

A set of national achievement examinations may actually work against school reform, argues a Berkeley researcher.

[In preparing this article, the author wishes to thank Alex McLeod, who directed the British end of the project, and Ellie O'Sullivan for her advice on British examination classes. She is also grateful to the teachers and students in both countries who participated in the project. Much of the substance of this article appears in Exchanging Writing, Exchanging Cultures: Lessons in School Reform from the United States and Great Britain, Cambridge, MA: Harvard University Press, 1994. Sections are reprinted with permission of the publisher.]

Warren Simmons and Lauren Resnick of the New Standards Project¹ present a wonderful vision for U.S. schools in the twenty-first century. They imagine schools in which students have "worked on extended

¹ The New Standards Project is based at the National Center on Education and the Economy and at the University of Pittsburgh's Learning Research and Development Center.



Sarah Warshauer Freedman is a professor of education at the University of California at Berkeley and director of the National Center for the Study of Writing and Literacy. She currently is working on a national teacher research project with English and social studies teachers who teach multicultural groups of students in urban schools.

projects, discussed complex problems, and generally thought their way through a demanding curriculum aimed at the kinds of knowledge and skills they will need as citizens and workers of the future.”

Significantly, Simmons and Resnick envision that these high expectations will apply equally to poor, minority, and immigrant children as well as to children in upscale suburban schools. If we can implement their vision for education, we will indeed produce a well-educated populace, ready to lead our country into the next century.

Framing the Problem

The question is how best to achieve this vision. Consistent with the current federal agenda, Simmons and Resnick suggest one approach, the creation of “world-class content and performance standards, a performance-based examination system that embodies those standards, together with...procedures for scoring students’ work reliably and fairly.” *America 2000*, the blueprint for educational reform produced by the National Governors Association in 1991 and spearheaded by then Governor Clinton, calls for voluntary, national, high-stakes achievement examinations in core subject areas for grades four, eight, and twelve. The scores from such tests would eventually be given to potential employers or used for college admission. High national standards and goals would be attached to the tests.

According to Simmons and Resnick, the goal is to build “a revitalized education system using assessment as a tool for transforming instruction and learning.” Marc Tucker, co-director with Resnick of New Standards, advances the argument for reform for such testing: “If as some thoughtful people have suggested recently, our schools are actually test-preparation organizations, then the current movement toward national standards and examinations may turn out to be the most powerful reform strategy we have.”

If his claim is true that our schools are “test-preparation organizations,” I contend that *that* needs to be reformed. Over the past several years I have been studying the effects that national exams have had on school reform in Great Britain. The evidence from the British experience suggests that a system of high-stakes examinations—even well-designed, performance-based examina-

tions—presents a flawed foundation on which to build a national educational reform movement. My evidence further suggests that high stakes exams have the potential to move us away from, rather than towards, the end point we all want to achieve.

A stronger foundation for educational reform is more likely to come about when teachers and school administrators rethink and then reshape the curriculum as well as the organization of the school. In this case testing would follow from—but not lead—the reform effort.

I have come to these conclusions after extensive observations of British secondary schools, where high stakes national examinations attached to high standards for teaching and learning have been in place since the beginning of this century (Freedman, 1994). It makes good sense to examine the British system, since U.S. proposals for national achievement examinations seem to have been inspired by, if not modeled on, the British system (Madaus and Kellaghan, 1991). Although one can never directly transpose findings from one culture to another, it is possible to learn much from others’ experiences.

Focus of Study

In my research of the British examinations and their effects on what and how students learn, I focused on just one area of the curriculum, English language and literature. Part of the study involved a national questionnaire of 695 teachers across grade levels, and 702 of their students at the secondary level in both the United States and Great Britain (England, Scotland, and Wales). It also included observations in eight English/language arts classrooms, four in the United States and four in Britain, across the equivalent of grades six to nine.

The teachers who completed questionnaires and the teachers of the eight case study classes were selected to represent the most thoughtful current practice in their respective countries. The case study classes were paired—one in England and one in the U.S.—and the students in the paired classes exchanged their writing with one another for an entire academic year.

For the writing exchanges, the two teachers in each pair coordinated their curricula so that students were doing roughly the same kinds of writing at the same time. Although students sent personal letters back and forth,

The teachers of the exam classes had to adhere to requirements that inhibited the amounts and kinds of writing students did.

the main focus of the exchange was on major and substantive pieces of writing—autobiographies, books about school and community life, opinion essays, and essays about literature.

Research teams in both countries observed the classrooms and interviewed students and teachers about the ongoing instructional process. Two of the four British classrooms that I studied intensively were ninth-grade equivalent and were in the first year of the two-year examination course for English language and literature.

The British Exam System

In Britain during the equivalent of grades nine and ten, the entire curriculum in every subject for those two years consists of an examination syllabus that prepares students for the General Certificate of Secondary Education (the GCSE). Students normally are taught by the same teacher for both years of an examination course. They usually take GCSE exams in six or more different subjects. For example, besides English language and literature, they take subjects such as mathematics, geography, drama, Spanish, French, Russian, biology, chemistry, and physics.

These exams determine each student's future. Those with good results (about 30% of the test takers) continue in school to prepare for university or other post-secondary education, or leave with a fair chance of employment. Those with poor results generally leave school and face bleak job prospects.

Except for those enrolled in a specific vocational program, students who continue in school take an additional two-year examination course. This prepares them for yet another set of exams, the A levels, which determine university admissions. Only about 20% of these students—or something less than 6% of the nation's 18-year-olds—actually complete the A level course and take the examinations.

During the writing exchanges in 1987-88, the British national examinations exemplified, in every way, what we in the United States would consider the most ideal examination system. The British system had just been revised. Previously students had been separated into two examination tracks: one for the university bound and the other for the non-university bound. However, beginning in 1987, all students enrolled in the same courses and took the same exams, the GCSE.

In addition, for the English language and literature examinations, schools were able to choose the option of coursework (a portfolio) as the only basis for evaluation. In each portfolio, the examining board required that at least one piece of writing be completed under "controlled" or testing conditions on an impromptu topic. But class work comprised the remainder of portfolio material. The students and their teacher selected writing for the portfolio together. The work was to represent each student's best efforts, and to demonstrate a range of knowledge and skill.

In exam portfolios that I collected, I found answers to questions on literature, imaginative writing related to literature, expository essays about literature and current events, and sometimes original fiction, autobiography, or other personal writing. In every sense, this was what we in the United States call performance assessment.

With respect to scoring, completed portfolios first were graded by a committee of teachers at the student's school, one of whom coordinated school marking with standards set by a national board. Then the portfolios were sent to the national board to be spot checked for consistency with national standards. In cases where inconsistencies were detected, the portfolios were scored again. Every student portfolio was graded as a whole; no grades were given for individual pieces.

Before the coursework-only option on the national exam, students had been evaluated solely by their performance on a "terminal examination" at the end of the two-year course. This terminal examination consisted of impromptu essay questions and writing prompts, given in a testing setting. In 1987, most British teachers were pleased with the coursework-only direction the GCSE examination was taking.

Curriculum Inhibitions

In spite of the new examination system, my research revealed that the national examinations were having a negative effect on teaching and learning in Britain. The exam classrooms were noticeably different from the U.S. classrooms I studied and the British classrooms for younger students (those who were not yet preparing for exams). The teachers of the exam classes had to adhere to requirements that inhibited their ability to build a coherent curriculum



with their students, and inhibited the amounts and kinds of writing the students did. The constraints of the examinations also colored the teachers' responses to their students' writing.

When I began working with my British colleagues to plan the writing exchanges and with Philippa Furlong and Gillian Hargrove who taught the British exam classes, we all assumed that coursework-only examinations, which required a variety of types of writing, would be entirely compatible with the goals of the writing exchange. We expected the teachers to merge writing for the exchange and writing for the examinations. However, the pressure of the examinations took over. The high stakes of the examiner audience and the teachers' and students' perceptions of the kind of writing that was appropriate for the exams constricted how and what students wrote. Philippa explained the inhibitions her students felt when writing for the examiner:

The kids didn't feel that confident about really baring their souls in an assignment that was going to go off to an examiner even though they knew that I was one of the examiners. They knew that beyond me there was an unknown quantity.

In the end, Philippa's students decided that most exam writing was a kind of writing that

was just for the examiner and would be of no interest to anyone else. They couldn't reconcile the examiner audience with any other audience, including the students in the U.S.

Across all eight classes involved in the writing exchanges, I found that students were most committed to their writing when they decided on what to write with their teacher. They were motivated through their own choice-making and by being part of a community working together. Students in the exam classes—unlike the two groups of younger British students who had no exams and unlike their U.S. partners—showed this kind of involvement only when they were not writing for the examiner, which was rarely.

When their students wrote for the exam audience, Gillian and Philippa inadvertently began to take away their students' responsibility over the subject matter for their writing. They acted contrary to their own theories about how students learn to write, which includes having students assume responsibility as a key component.

Besides the pressures of the high stakes and the heavy workload across all their subjects, the examination restricted the teachers in how they could help their students. For the 1994 exam syllabus, the British testing board issued the following policy guidelines about how much help teachers could give students on their writing:

In the British exam classes, rules restricted how much writing help teachers could give students.

Advice remains on a general level, only becoming specific to exemplify general comments, and the onus is left on the candidate to incorporate the teacher's general advice by making specific alterations and thus submitting a final draft. This level of advice is acceptable whereas proof reading, in which the teacher points out a detailed series of errors, omissions, and amendments for the candidate to correct in the subsequent draft, is inadmissible.

Such policies inhibited British teachers from the kind of explicit teaching needed to reach the most needy students. These requirements are especially problematic when teaching writing to bilingual students, who often learn from specific corrections and by having errors explained to them in some detail. In this context, it is crucial to remember that most students score too low to continue on the next exam course that would prepare them for university entrance.

British education policies also appear to have affected students' motivations for writing. The reaction of a student named Surge was typical. For him, writing in school and not for the exchange had only one value—it helped him prepare for the GCSE. By contrast, the exchange audience was real and important to him: "It makes a lot of difference to me that someone in America is reading my work, to think that it comes all the way from England."

Another student, Leabow, explained how hard she worked to connect with the exchange audience, which is something she did not do for the GCSE examiner. "You had to...build up a personality...so they [the U.S. students] could imagine what you were like through your personality...[and] build up an image for yourself to make them see...who you are."

Implications for U.S. Educators

Currently the British examination system has become more conservative and many teachers feel it is deteriorating. It now includes more emphasis on terminal exams and less emphasis on portfolios or coursework. (The British government feared that by using only portfolios, it was difficult to know what students could do independently.) Ultimately, British teachers perceived issues such as good curriculum, good teaching, and high standards for student

performance as being unrelated to the national tests or to national standard-setting efforts. The U.S. teachers left the experience convinced that any kind of high-stakes examinations, when tied to the curriculum, would be harmful to their students' writing development.

What happened in the British exam classrooms suggests that the path to curricular reform through examinations, though tempting, remains elusive. As U.S. policy makers and educators contemplate high-stakes examinations, they remain optimistic that the "right" kind of examinations will lead to improved instruction. Exams are popular among policy makers because they provide one of the few levers on the curriculum that they can control.

And there are some potentially positive side effects, too. When teachers participate in the reform process, it engages them in substantive professional debates about standards and standard setting. It also contributes to their professional development as they create and score performance-based exams.

However, when exams take control of something as personal as writing, the teacher and students no longer work together to own the writing. Rather, the writing is owned by a distant examiner. In the British exam classes, individual pieces of writing were rushed; extended pieces that were common in the early years disappeared. Rules restricted teachers with respect to how much help they could give their students.

An exam system can affect curriculum negatively, especially when the stakes are high and the exam is used to sort students in ways that determine access to higher education and to valued job opportunities. When exams function in this way, the pressures on the classroom become formidable and not necessarily positive.

Professor Linda Darling-Hammond of Teachers College argues that any exams that function to sort people, discourage educational equity and fail to promote school reform (1994).

Changes in the forms of assessment [such as by using portfolios and performance assessments] are unlikely to enhance equity unless we change the ways in which assessments are used as well: from sorting mechanisms to diagnostic supports; from external monitors of performance to

locally generated tools for inquiring deeply into teaching and learning; and from purveyors of sanctions for those already underserved, to levers for equalizing resources and enhancing learning opportunities.

Darling-Hammond suggests that assessments will only be useful reform tools if they function as "top down support of bottom up reform." Similarly, George Madaus, a professor at Boston College, concludes that "the nation cannot test its way out of its educational problems ... it is the teachers, not tests or assessments, that must be the cornerstone of reform efforts" (1993).

These arguments are consistent with those who argue that reformers must begin by working collaboratively with schools and the communities they serve, involving teachers, administrators, students, and parents. Examples of well-known reform efforts in this vein include James Comer's School Development Program, Henry Levin's Accelerated Schools, and TheodoreSizer's Coalition of Essential Schools. Granted, the reach of these efforts remains limited, and none attends equally well to every aspect of the problem. More importantly, none can begin to solve the larger social problems that plague schools serving high percentages of students from conditions of poverty, such as inadequate health care, poor nutrition, and lack of economic opportunity. Still, these reformers are taking sensible first steps. They offer a welcome counterpoint to those who advocate reform-via-exam. □

[The research for this article was supported under the Educational Research and Development Center Program for the National Center for the Study of Writing, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed here do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.]

References

- America 2000: An Education Strategy Sourcebook* (ED/OS91-13). Washington, DC: U.S. Department of Education, 1991.
- Cheney, L. *National Tests: What Other Countries Expect Their Students to Know*. Washington, DC: National Endowment for the Humanities, 1991.
- Darling-Hammond, L. "Performance-Based Assessment and Educational Equity." *Harvard Educational Review* 64 (1): 5-30 (1994).
- Freedman, S. *Exchanging Writing, Exchanging Cultures: Lessons in School Reform From the United States and Great Britain*. Cambridge, MA: Harvard University Press, 1994.
- Madaus, G. "A National Testing System: Manna From Above? An Historical/Technological Perspective." *Educational Assessment* 1 (1): 9-26 (1993).
- Madaus, G., and Kellaghan, T. *Examination Systems in the European Community: Implications for a National Examination System in the United States*. Paper prepared for the Science, Education, and Transportation Program, Office of Technology Assessment, U.S. Congress, Washington, DC., 1991.
- National Council on Education Standards and Testing. *Raising Standards for American Education: A Report to Congress, the Secretary of Education, the National Education Goals Panel, and the American People*. Washington, DC: National Council on Education Standards and Testing, 1992.
- Simmons, W., and Resnick, L. "Assessment as the Catalyst of School Reform." *Educational Leadership* 50 (5): 11-15 (February 1993).
- Statistics of Schools, January 1988*. London: Department of Education and Science, 1989.
- Tucker, M. "Toward National Standards: Measuring Up." *America's Agenda* 21-22 (Fall 1992).