

## **US Perspectives on Student Assessment in Other Countries**

Report Issued by the US-China Center for Research on Educational Excellence  
Prepared by David E. Kirkland

In the US, student assessment has become the engine driving education, from classroom reform to classroom practice, from education research to federal policy. According to proponents, student assessment is the means to educational improvement, to tell whether or not our system of education is working and/or what parts are not working so that they may be fixed or abolished, ensuring all children a fair, decent, and quality education (see the No Child Left Behind Act of 2001 [NCLB]). However, given our increased reliance on assessment, specifically student assessment, in the wake of NCLB and the growing paranoia surrounding our nations “crippled” education system, there has been a movement to emulate foreign assessment-based models of education in our own national education system (again, see NCLB; Hirsch, 1996). While we stand to learn much from how other countries do education, basing our education system on student assessment alone might cost us far more than it might give us. Further, US researchers who study student assessment in countries other than the US paint a rather cautioning picture, which deserves our attention.

In this report, I review significant literature which together document how US researchers perceive student assessment beyond US borders. I use the term “assessment” to identify measurements of student learning and performance, begging together different types of assessment instruments, including achievement tests, minimum competency tests, developmental screening tests, aptitude tests, observation instruments, performance tasks, and authentic assessments. As such, the review defines assessment primarily in relation to tests used to interpret students learning as it reflects the overall quality of schooling in a given country. First, I briefly describe the history of student assessment and talk about how student assessment is used in other developed countries. Lastly, I pull together the sentiments of US researchers interested in student assessment in other countries and tell what their thoughts might mean for us.

### **A Brief History of Student Assessment**

The mass formal assessment of students is a relatively modern phenomenon. Morgan (2001) shows in her study of the content and institutions of Roman education that widely standardized practices could function, without formal examinations, by means of high levels of competition and other less formal means of assessment. For Morgan, the role of educational competition defined and maintained social elites in the Roman Empire. Following Morgan, Kehm (2001) explains how oral examinations become one of the first forms of formal assessment. In 18<sup>th</sup> century England, Kehm writes, the technology of writing replaced speaking as the preferred method for testing. This was a radical shift, as a century prior, examinations were conducted in public, orally and in Latin, with the participation of the academic community.

Formal written assessment accompanied the world on the eve of modernity. Societies were gearing themselves to become more efficient, more scientifically effective. Stray (2001) gives an account of the shift from oral to written examinations at Oxford and Cambridge universities in the 18<sup>th</sup> and 19<sup>th</sup> centuries. In this study, Stray outlines four factors as being crucial to this shift: the move from group socio-moral to individual cognitive assessment in the later 18<sup>th</sup> century; the differential difficulty of oral testing in different subjects; the impact of increased student numbers; and the internal politics of both Oxford and Cambridge.

While oral assessments retained some value, formal written examinations gained wide use by the early 20<sup>th</sup> century. According to Sutherland (2001), occupational groups aspiring to legitimate professional status devised their own examinations. Subsequently, these groups also came to use the examinations of an expanding education system. In this way, we begin to

understand the ways in which assessment functioned to legitimate the elite of society. In deploying the instrument of examination, professions as well as the academy were also invoking the challenge to patronage and the appeal to merit which it symbolized (Boweles & Gintis, 1976). Yet, as Sutherland points out, such use did not cut professionals loose from the class structure and status hierarchy of the society. Rather, it anchored them firmly in a powerful position within it. According to Black (2001), systems of assessment are shaped in part by personalities and institutions who pursue value and social improvement.

The modernist notion that student assessment can help a society determine (or at least quantifies/measures) its potential is a popular belief. From the exercise of IQ tests to the “dream” (Black, 2001) that standards (both social and educational) can be raised by external testing greatly influences the growing use of student assessment worldwide. Kellaghan (2001) describes in his study of the globalization of assessment in the 20<sup>th</sup> century the considerable increase in the 1990s in the number of countries involved in national and international assessments of pupils in their education systems. While participation in international comparative studies is largely confined to industrialized countries, many developing countries have even begun to carry out national student assessment in the last ten years. Growth in the use of student assessment at the system level in the context of factors associated with globalization (e.g., a utilitarian view of education, the need to develop human capital, concern for performance/results and the use of standard procedures and advanced technology) should, in fact, give rise to concern.

### **By the Tape of Others: Why No Country Will Ever Measure Up**

On December 7, 2004, the Wall Street Journal headlined a story entitled, “Economic Time Bomb: US Teens are Among Worst at Math.” The story came on the heels of the disheartening results of the latest international assessment of student achievement. The results of the Program for International Assessment (PISA), an ongoing study which features 15-year-olds from 41 countries, ranked the US 24<sup>th</sup> out of the 29 member nations belonging to the Organization for Economic Cooperation and Development (OECD). However, the Trends in International Mathematics and Science Study (TIMSS), which evaluates students in the fourth and eighth grades, told a somewhat different story. According to TIMSS (2004), US student performance had in fact improved in the four years since the previous assessment. Both studies attest to two things: assessment of student performance is a growing trend in international comparative education studies and the assessment trend is fundamentally marked by controversy and uncertainty.

We should not be so concerned that the results of the TIMSS and PISA do not match. The two tests serve different political and ideological interests. TIMSS examines student performance and the background characteristics of students, teachers, and schools. Assessment items, which are developed through a consensus of representatives to the International Association for the Evaluation of Education Achievement (IEA), are designed to link directly to the curricula of the participating countries. The TIMSS report thus specifies what students are expected to learn and how well they are learning it. For the 2003 assessment, IEA revised the TIMSS framework to incorporate the curricular changes in participating countries, so that it “accurately” reflects what educators in all these countries are currently including in their math and science curricula.

PISA, which developed under the OECD, has a different purpose. It measures literacy in reading, mathematics, and science in 15-year-olds. For example, in mathematics, PISA assesses how well 15-year-olds can recognize and interpret mathematical problems in their world, translate problems into mathematical context, and use mathematical knowledge and procedures to solve problems. PISA is not directly tied to the school curriculum but was conceived and designed to assess the practical outcomes of education systems. In other words, the PISA

assessment aims to determine whether students not only have the knowledge they need but also the ability to use it to solve real problems.

Tests are administered to 15-year-olds because that is typically the last year of compulsory schooling in participating countries. The most recent PISA emphasized math and also added a separate section independent of any particular content area devoted to problem solving. Hence, TIMSS and PISA offer different assessments, which measure different things. Given the hysteria surrounding the results of TIMSS and PISA and the assessment driven mandates of NCLB, we should be concerned with the growing and popular tendency of educational organization use of student assessments to measure the quality of national and international education systems.

### **Student Assessment beyond Our Borders**

US researchers have revealed that student assessment do not match in any two countries. In England, Silvermail (1996) found that schools provide multiple measures of performance that serve multiple purposes. Schools assess student progress through both a national examination and teacher-made tests when students reach the ages of seven, eleven, fourteen, and sixteen. These exams are primarily used to measure the effectiveness of the schools in delivering the national curriculum. All schools must follow nationally prescribed content and pedagogic methods and set targets for individual pupil learning (Whetton, Twist, and Sainsbury, 2000). Gipps, Clarke, and McCallum (1998) assert that this system places too little emphasis on assessment of learning and too much on assessment for learning.

The French educational system employs three sets of national assessments of student progress. At the beginning of grades three, six, and nine, students are tested for diagnostic and planning purposes. By the end of grade nine, students are tested to measure attainment. At the end of high school, tests determine students' professional and educational futures. Researchers point out that although French educators complain about the complexity of the high-stakes, end-of-high-school exam, the public and the media support it (Fowler, 2001).

In 1999, Cheng investigated national student assessments used in Hong Kong to assign students to primary and secondary school programs. Cheng discovered that such assessments increasingly dictate classroom instruction. According to Cheng, Hong Kong's education leaders are aware that "if it is not examined, it won't be taught" (p. 23). In a similar vein, Feng (1999) reports that both the Chinese government and the Chinese public consider the National College Entrance Examination (NCEE) as crucial for China's political, economic, and educational development. As such, the Chinese poor have increasingly come to depend on the NCEE to equalize their opportunities for an elite education and social future. Students in Japan apply to the high schools that they want to attend and get admitted on the basis of their performance on the schools high-stakes examinations. The only national examination in Japan's public education system is the college entrance examination conducted by the Ministry of Education every January (Posthletwaite, 1995). Students with high scores on the common national exam must then take individual university exams at the schools that they wish to attend. The time that Japanese adolescents devote to high school and university examination preparation—attending special cram schools, studying with tutors, taking repeated mock exams, and studying late into the evening—is referred to in a report released by the US Department of Education (1998) as "exam hell."

Given the inconsistency in large scale, standardized student assessments, it is impossible to use national examinations to compare student achievement among different countries (Holloway, 2003). As such, there is no uniform approach to student assessment in different nations. Further, the US is the only country in the developed world without a national curriculum, national standards, or national assessment (Bracey, 1994). Based on data

haphazardly taken form studies like TIMSS and PISA, some researchers are beginning to argue that this is a good thing, while many others maintain it is not. Those in favor of national student assessment argue that “[other] countries are educating everybody to higher levels than we are, and that the main differences between us and them is standards: they have them and we don’t” (Spillane, Robert-education week).

### **What’s Going On**

While some studies suggest that US students are indeed lagging behind much of the developed world, an article by Michael Feuer of the National Research Council and Kathleen Fulton of the Office of Technology Assessment (1994) sheds some light on what the situation actually is in the US with respect to the rest of the world. What’s really going on?

For one, the authors show that some countries are moving away from the national assessment model of education even as the US is etching closer to it. In addition, they point out that, while countries outside the US have a long history of limiting access to upper levels of education, as equity issues have gained worldwide attention in recent decades, the “gatekeeper” function of student assessment has come under considerable international fire. In this way, their study offers a stunning glimpse at the unintended consequences associated with assessment driven education systems. For example, over 60% of US students enter into higher education. In France on the other hand, by age 15, only 67% of students attend schools leading to the baccalauréat, which is needed for college admissions. Of these, only 50% pass the exam and earn the baccalauréat, which means only 38.5% of France’s students are even eligible to attend college.

Many US educators sulk in admiration of the German Arbitur examination and have argued that the U.S. should emulate it. In spite of its popularity, the German Arbitur is not a national examination, nor is it even a single form of an examination (Bracey, 1994). Further, given its well-regarded reputation, the Arbitur itself sets up social barriers. In 1986, “only 24% of the relevant age group had passed the exam” (Bracey, 1994, p. 167). While those who pass the Arbitur have a constitutional right to attend a university in a chosen field of study, recent increases in the number of eligible students have caused some restrictions on this right. According to Bracey, it is determined early, by age 10, that only a small percentage of German students will pass the Arbitur. After age 10, some students are sent to lower general education for another five years, followed by terminal vocational training. Others are assigned to general education for six years, which prepares them for intermediate positions or occupations.

Bracey acknowledges that, in contrast to France and Germany (and to Japan and China), which use assessments as a fine seine to determine who goes on to higher education, Sweden does not use exams for college admission at all. For democratic reasons, Sweden determined that all of its citizens receive the same education through age 16 yet has a highly prescriptive and highly centralized curriculum. After age 16, Swedish students can enroll in upper-secondary school, which offers two-, three-, and four-year programs. Admission to the upper-secondary programs is determined on the basis of teacher grades, not tests. (Work experience used to be a large consideration also.) The main use of standardized testing, which is voluntary through grade 7, limited to English in grade 8, and limited to Swedish and mathematics in grade 9, is to let teachers see how their students are performing compared to the nation as a whole. Teachers then use this information to adjust their grading standards.

Feuer and Fulton (1994) point out that in countries in the European Union, “the standardized test has all but disappeared before age 16” (p. 33). They note also that, if the U.S. imposes a common test for fourth-graders, it would be the only industrialized nation in the world to have such an assessment. Drawing lessons for America from the European experience with testing, Feuer and Fulton argue that European nations look to the National Assessment of Educational Progress (NAEP) as a model because of its sampling methodology, its wide range of

skills tested, and its low-stakes nature. The lesson is to be cautious in making changes in the NAEP. Second, Europeans don't make much use of tests to evaluate the quality of schools. The lesson is that we should pay more attention to other indicators of quality. Third, Japan and France typically outscore the U.S. in international comparisons. Yet Japan uses mostly multiple-choice tests, which France does not use at all. "Changes in test format do not automatically lead to better assessments of student achievement, to more appropriate uses of tests, or to improvements in academic performance," write the authors (Feuer & Fulton, 1994, p. 37).

Feuer and Fulton observe that, because the U.S. has 15,000 autonomous school districts, national agreement on the contents of the curriculum is much harder to attain here, usually resulting in some least-common-denominator tests of minimum basic skills. Even the NAEP does not assess skills and content in the complex ways found in the other national assessments. Finally, Feuer and Fulton point out that teachers in a number of countries have considerable responsibility for the scoring of examinations. They contend that "faith in the professional caliber of teachers is a necessary condition for a credible system of examinations that requires teachers' judgments in scoring" (p. 37). By implication, such faith is not present here. In the same vein the researchers note that many other countries have only a few institutions that train teachers—sometimes only one—a fact that makes it easier to reach consensus about what is to be taught and how.

## **Conclusion**

While student assessment does not seem to be going away anytime soon, we must be cautious of how these instruments can be used. If used correctly, research suggests that they can point educators in the right direction with respect to student needs. However, if used as a weapon in international and national educational warfare, researchers also warn that such instruments might yield mass destruction. Instead of trying to compare the US to the world or vice versa, we need to move our education system toward the preparing student for the promising idea of democracy, to fulfill the fundamental tenets by which the western world is now established. This is the tape by which we must measure others and ourselves. As it stands, student assessment has been used as an instrument to undermine both education and democracy, to erect elite barriers which distinguish groups of people based on unfair and arbitrary outcomes. Indeed, our nation is at risk. And it is not because our students or students in other countries are failing. Our nation is at risk because we are failing the greatest test of all: to write a system of education in which no one fails.

## REFERENCES

- Black, P. 2001. Dreams, strategies, and systems: Portraits of assessment past present, and future. *Assessment in education: Principles, policy and practices*, 8 (1), pp. 65-85.
- Bracey, G. W. 1994. Standards and Assessment. *Phi Delta Kappan*, 76, p. 166-7.
- Bybee, R. W. 2005. No Country Left Behind. *Issues in Science and Technology*, 21 (2), p. 69-75.
- Cheng, L. 1999. Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253-271.
- Feng, Y. 1999. National college entrance examinations: The dynamics of political centralism in China's elite education. *Boston University Journal of Education*, 181(1), 39-57.
- Feuer, M.J., & Fulton, K. 1994. Educational Testing Abroad and Lessons for the United States. *Educational Measurement: Issues and Practice* 13 (2) (1994): 31—39.
- Fowler, F. (2001). Testing French style. *The Clearing House*, 74(4), 197-200.
- Gipps, C., Clarke, S., & McCallum, B. 1998, April. The role of teachers in national assessment in England. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California (ERIC Document Reproduction Service No. ED 419 836).
- Gonzales, P., Guzman, J., Partelow, L., Pahlke, E., Miller, D., Jocelyn, L., Kastberg, D., & Williams, T. 2004. *Pursing Excellence: Fourth grade Mathematics and Science Achievement in the United States and Other Countries from the Trends in International Mathematics and Science Study (TIMSS) 2003*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics (NCES).
- Holloway, J. H. 2003. A Global Perspective on Student Accountability. *Educational Leadership*, 60 (5), p. 74, 76.
- International Study Center, Lynch School of Education, Boston College. 1999. A bridge to school improvement: 1999 TIMSS benchmarking, executive summary [Online]. Available: [http://isc.bc.edu/timss1999b/pdf/TB99\\_Math\\_ExSum.pdf](http://isc.bc.edu/timss1999b/pdf/TB99_Math_ExSum.pdf).
- Kellaghan, T. 2001. The globalization of assessment in the 20<sup>th</sup> Century. *Assessment in education: Principles, policy and practices*, 8 (1), pp. 87-102.
- L. Ma. 1999. *Knowing and Teaching Elementary Mathematics*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Lemke, M., Sen, A., Partelow, L., Miller, D., Williams, T., Kastberg, D., & Jocelyn, L. 2004. *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results from the U.S. Perspective*. Washington, D.C.: U.S. Department of Education, NCES.
- OECD Programme for International Student Assessment. (n.d.). Summary of first results from PISA [Online]. Available: [www.pisa.oecd.org/knowledge/summary/intro](http://www.pisa.oecd.org/knowledge/summary/intro).

htm.

Postlethwaite, T. (Ed.). 1995. *International encyclopedia of national systems of education (2nd ed.)*. Tarrytown, NY: Pergamon.

Silvernail, D. 1996. The impact of England's national curriculum and assessment system on classroom practice: Potential lessons for American reformers. *Educational Policy*, 10(1), 46-62.

Stray, C. 2001. The shift from oral to written examination: Cambridge and Oxford 1700-1900. *Assessment in education: Principles, policy and practices*, 8 (1), pp. 33-50.

Sutherland, G. 2001. Examinations and the construction of professional identity: A case study of England 1800-1950. *Assessment in education: Principles, policy and practices*, 8 (1), pp. 51-64.

U.S. Department of Education, Office of Educational Research and Improvement. 1998. *The educational system in Japan: Case study findings*. Washington, DC: Author.

Whetton, C., Twist, E., & Sainsbury, M. 2000. National tests and target setting: Maintaining consistent standards. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, Louisiana (ERIC Document Reproduction Service No. ED 441 849).