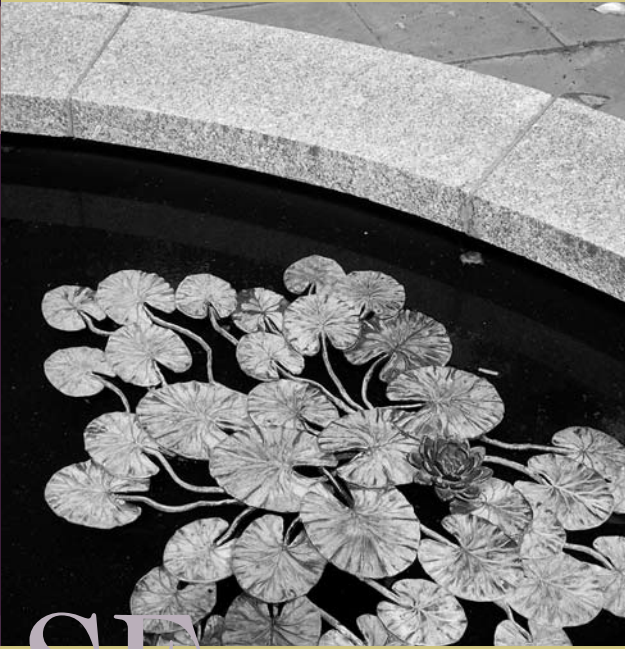


@Penn
GSE
@PennGSE



A Review of Research

Is High-Stakes Testing Working?

By Jonathan Supovitz

Test-based accountability—the use of tests to hold individuals or institutions responsible for performance through the application of rewards and/or sanctions—has become the cornerstone of U.S. federal education policy, and the past decade has witnessed a widespread adoption of test-based accountability systems in states throughout the country. Consider just one material manifestation of this burgeoning trend: test sales have grown in constant dollars from approximately \$260 million annually in 1997 to approximately \$700 million today—nearly a threefold increase.

What influence has our substantial investment in testing and test-based accountability policy had on the behavior and performance of

MORE INSIDE...

Updates on Research and Publications from the Penn GSE faculty

A publication of the University of Pennsylvania Graduate School of Education (Penn GSE)

www.gse.upenn.edu/review

Continued on page 3

on the cover:
Lily Composition #3, Anne Froehling,
Steinhart Plaza, Bronze, 1993

@ PennGSE

A Review of Research

is produced three times annually
by the University of Pennsylvania
Graduate School of Education,
3700 Walnut Street
Philadelphia, PA 19104-6216
editor@gse.upenn.edu

Reproduction of these articles
requires the written permission
of the editor. ©2010 by the
Trustees of the University of
Pennsylvania. Please contact
the editor for references.

Tom Kecskemethy
Director, Communications

Nancy Brokaw
Editor

Contributing Writers
Liza Hill
Wayne Kobylinski

Nondiscrimination Statement
The University of Pennsylvania values diversity and seeks talented students, faculty and staff from diverse backgrounds. The University of Pennsylvania does not discriminate on the basis of race, sex, sexual orientation, gender identity, religion, color, national or ethnic origin, age, disability, or status as a Vietnam Era Veteran or disabled veteran in the administration of educational policies, programs or activities; admission policies; scholarship and loan awards; athletic, or other University administered programs or employment. Questions or complaints regarding this policy should be directed to: Executive Director, Office of Affirmative Action and Equal Opportunity Programs, 3600 Chestnut Street, Sansom Place East, Suite 228, Philadelphia, PA 19104-6106 or 215 898-6993 (Voice) or 215 898-7803 (TDD).



Is High-Stakes Testing Working?

By Jonathan Supovitz

Continued from page 1

American educators and students? Is high-stakes testing a substantive reform, or an intervention that merely reveals shortcomings in the system but does little to actually improve teaching and learning? This article summarizes a review of the evidence from the past decade of test-based reform (Supovitz, 2009).

Why We Test

Four major theories underlie our current reliance on high-stakes tests: *motivational theory*, which argues that test-based accountability can motivate improvement; the *theory of alignment*, which contends that test-based accountability can spur alignment of major components of the educational system; *information theory*, which posits that such systems provide information that can be used to guide improvement; and *symbolism*, which maintains that testing systems signal important values to stakeholders.

Motivational theory is the predominant theory underlying test-based accountability. According to this concept, the extrinsic rewards and sanctions associated

with the high-stakes test serve to motivate educators across the system to improve their performance. This presumes that lack of will, not lack of capacity, stands in the way of system-wide improvement.

The *theory of alignment* holds that system-wide improvement is most likely to occur if educational leaders align the major components of the educational system that surrounds schools (standards, curriculum, and assessments) so that they mutually reinforce each other. Alignment is usually thought of in terms of synchronizing the surrounding system, but can also be thought of as coherence between the external accountability system for schools and schools' sense of internal accountability (Ablemann et al. 2004).

Information theory maintains that student performance data are useful for teachers and administrators to make decisions about students and programs and that providing such data to local educators and giving them incentives to improve their performance will guide classroom and organizational decision-making.

Symbolism theory has also contributed to the growth and prevalence of high-stakes testing. In this model, the accountability system is seen to signal important values to stakeholders and, in particular, the public (Airasian 1988). This particular theory is manifested in the notion of "public answerability"—the idea that the public has a right to expect that its resources will be used responsibly and that public institutions are accountable to the public. High-stakes assessment results thus serve as evidence that public education is, in essence, responsible and rigorous.

Is high-stakes testing a substantive reform, or an intervention that merely reveals shortcomings but does little to improve teaching and learning?

The Movement Toward Measuring Outcomes

Over the past 20 years, the nation has shifted from the tracking of educational inputs (e.g., per pupil expenditures, teacher salaries, class size, required courses, seat time) as indicators of educational performance to an increased emphasis on testing as a means to hold schools accountable for educational outcomes. Standardized test results have become the primary indicator of school and student performance, with public reporting, monetary or nonmonetary rewards, a range of interventions for low-performing schools, and even state takeover, as consequences for the quality of performance (Elmore et al. 1996; Fuhrman and Elmore 2004).

By the early 1990s, standardized, multiple-choice high-stakes testing came under siege from many constituencies for containing gender bias, ethnic prejudice, and socioeconomic favoritism. Critics bemoaned the narrowing of curriculum and instruction and the perverse incentives inherent within high-stakes testing to retain and reclassify students. Many maintained that multiple-choice testing, with its emphasis on recall of isolated bits of knowledge, represented an outdated behaviorist view of learning. Moreover, research confirmed many of these critiques (see, for example, Garcia and Pearson 1994; Supovitz and Brennan 1997; Hamilton 2003).

To address these problems, educators introduced a bevy of alternative forms of assessment (e.g., portfolios, performance assessments, and open-ended tasks). Advocates saw these as more valid measures of student performance and as potential catalysts for school reform. As several states and national organizations began incorporating alternative forms of assessments into their test-based accountability systems, researchers examined the influence of these new forms of assessments on policy and practice. Findings on the potential for alternative assessments to deliver richer, less biased measures of student performance, however, were mixed. Scoring reliability was found to be high in science performance assessments (Baxter et al. 1992) but unreliable in portfolio assessments (Koretz et al. 1994). Portfolio assessments were found to reduce racial/ethnic gaps in performance but exacerbate gender differences (Supovitz and Brennan 1997). Performance task content continued to produce gender-related biases (Jovanovic et al. 1994). Alternative assessments were also found to be cost prohibitive. For example, the cost of large-scale science performance assessments in California was estimated to be 20 to 60 times higher than standardized multiple-choice assessments for an equally reliable score (Stecher and Klein 1997). Research also indicated that teachers continued to organize instruction around the timing of high-stakes assessments, regardless of their format (Borko and Elliott 1999). Performance assessments also influenced curricular activities and assessment practices (Lane et al. 1999), and teachers still prepared students for the test rather than the larger learning goals of the curriculum (Stecher and Barron 1999). While collective research dampened the enthusiasm for alternative assessments, some elements were incorporated into the dominant forms of high-stakes testing, including open-ended writing and performance tasks.

The NCLB Era

In 2001, test-based accountability was incorporated into the No Child Left Behind Act (NCLB), a major federal reform intended to bring about widespread improvements in student performance and reduce inequities between ethnic groups and other traditionally under-served populations. NCLB required states to adopt test-based accountability systems, testing annually in reading, math, and eventually science from grades 3 through 8 and one year of high school. States were to define proficiency and adequate yearly progress (AYP) to get all students to proficiency in 12 years. Schools that failed to make AYP for two consecutive years would be identified for improvement and students from those schools would have the right to transfer to another public school. The legislation also required measurable objectives for sub-groups of students and for states to certify teachers as highly qualified.

Studies and analyses on NCLB have begun to emerge. A four-year analysis conducted by the Center for Education Policy (Renter et al. 2006) surveyed state policymakers, district administrators, and schools. Many of the study's respondents credited NCLB for rising student performance. At the same time, they indicated a narrowing of curriculum with a focus on reading and math that has reduced instructional time for other subjects, a shift reflecting efforts to align curriculum and instruction with assessments.

Likewise, a synthesis of pre- and post-NCLB literature (Herman 2004) concluded that accountability attracts teachers' attention to the extent that teachers are more influenced by testing than by the standards themselves. In practice, Herman found, test preparation merges with instruction, with a concurrent de-emphasis of non-tested content.

What about Interim Assessments?

Drawing on extensive interview and observation data from nine schools in two districts, a recent study focused on teachers' use of interim assessment data in elementary school mathematics. Findings showed that assessments linked to the curriculum did help teachers decide what content to re-teach and to whom. This practice was facilitated by school and district expectations that assessment results would be used to inform instruction, a high-quality information management system, time to re-teach content, instructional supports for students, and professional supports for teachers.

The evidence was limited, however, that interim assessments changed how teachers approached instruction. Changes to teachers' instructional practices depended on their capacity to interpret assessment data and their knowledge of mathematics for teaching.

From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction, by Penn GSE's Margaret Goertz, Leslie Nabors Oláh, and Matthew Riggan, is a publication of the Consortium for Policy Research in Education.

Some evidence suggests improvements in national performance associated with test-based accountability (Hanushek and Raymond 2004; Kober et al. 2008). But credit for these gains has also been given to school district policies and programs (Renter et al. 2006). Thus, while performance is improving, the contribution of high-stakes testing remains unclear.

Still other research has explored ways to use data from high-stakes tests to improve instruction—and again the verdict on the value of these assessments is mixed. These studies have typically found that the data provided by high-stakes exams contain general information about student performance but lack the nuance to provide fine-grained instructional guidance (Supovitz and Klein 2003). In response to such analyses, many districts have moved to more frequent quarterly or benchmark assessments (Goertz, Nabors Oláh, and Riggan 2009), but these instruments may suffer from the same problems.

What Have We Learned?

So, what has the past decade of testing policy taught us about the likelihood of improving the education system through high-stakes testing? Let's review the effects according to the four theories of testing policy described earlier:

1) *High-stakes testing does motivate educators, but responses are often superficial.* In the best cases, high-stakes testing has focused instruction toward important and developmentally appropriate literacy and numeracy skills—but at the expense of a narrower curricular experience for students and a steadier diet of test preparation activities in classrooms, particularly in low-performing schools, which are the targets of test-based accountability.

2) *Test-based accountability fosters alignment of the central components of the educational system.* The evidence does indicate that high-stakes testing encourages educators to align curriculum, standards, and assessments. However, it is uncertain whether narrower exams are being aligned to more comprehensive standards, or vice versa.

3) *High-stakes testing regimes have limits as information tools.* The data from high-stakes tests are useful to policymakers for assessing school and system-level performance but insufficient for individual-level accountability and provide meager information to teachers for instructional guidance.

4) *Test-based accountability is an appealing political strategy.* High-stakes testing answers a real need for policymakers to demonstrate to the public that they are spending tax dollars judiciously.

In sum, the evidence indicates that high-stakes assessments foster alignment in the system and provide an important representation of policymakers' commitment to improving the public education system. At the classroom level, high-stakes tests motivate teachers to change their practices, but these changes tend to be more superficial adjustments in content coverage and test preparation practices rather than engendering deeper improvements in instructional efforts. In addition, the data provided by annual assessments are of limited utility to teachers.

Where Do We Go From Here?

Over the past couple of decades, test-based reform in the United States has gone through two major cycles: first, a widespread exploration into a variety of alternative assessment forms, then an increased emphasis on annual testing and state test performance as the authoritative indicator of the quality of schools and districts. Despite their substantive differences, these two cycles were each born of the best intentions: a desire to raise the performance of students and to redress inequalities in the educational system. But these disparities are driven by our social priorities, not our educational system, and they will not be remedied by rewriting the tests.

Our experiences with high-stakes testing should teach us about their serious limitations. Rather than investing in substantial efforts to improve teaching and

While these tests certainly have a place in our efforts to improve our schools, their inflated role is indicative of our lack of will to enact deeper, more substantive reforms.

learning, we have developed a system that relies on summative testing as the cure for what ails us. While these tests certainly have a place in our efforts to improve our schools, their inflated role is indicative of our lack of

will to enact deeper, more substantive reforms. In short, we have been effective in setting goals and motivating educators through high-stakes testing—but we have done so without building their capacity to achieve the goals for which we are holding them accountable.

In the next decade, developments in two test-related areas may well contribute to how much substantive progress we make. First, teachers need better testing tools and the capacity to utilize them effectively. Increasingly, we have the expertise to embed valuable information about patterns of student understanding into assessments that can be used by teachers for real instructional guidance. With better information about student subject matter (mis)conceptions and problem-solving strategies, and the skills to act on that information, teachers can hone their instructional responses to improve students' understanding.

Second, we need to reform the reform. We must find a way to assimilate short-, medium-, and long-cycle assessments into a more coherent system that takes advantage of the strengths of each and ameliorates the undue influence that a single high-stakes assessment carries. A more robust assessment system might begin in the schools with more formative assessments, continue with a set of curriculum-specific interim assessments that act like guideposts, and culminate in a summative annual assessment. Such an aligned assessment system would reduce our emphasis on, and attention to, annual high-stakes assessments.

While technological advances make the integration and standardization of such concepts more feasible than ever before, the political challenges are considerable. But if we are serious about improving the education our youth receive, we must relegate high-stakes accountability to its proper place as a measurement and incentive companion to deeper instructional reforms.

Jonathan Supovitz is an associate professor at Penn GSE and a senior researcher for the Consortium for Policy Research in Education.

REFERENCES

- Ablemann, C.H., Elmore, R.F., Even, J., Kenyon, S., & Marshall, J. (2004). When accountability knocks, will anyone answer? In R.F. Elmore (Ed.), *School Reform from the Inside Out*, 133-199. Cambridge, MA: Harvard Education Press.
- Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high stakes testing. *Educational Evaluation and Policy Analysis*, 10(4), 301-313.
- Baxter, G.P., Shavelson, R.J., Goldman, S.R., & Pine, J. (1992). Evaluation of procedure-based scoring on hands-on science assessment. *Journal of Educational Measurement*, 29(1), 1-17.
- Borko, H., & Elliott, R. (1999). Hands-on pedagogy versus hands-off accountability: Tensions between competing commitments for exemplary mathematics teachers in Kentucky. *Phi Delta Kappan*, 80(5), 394-400.
- Elmore, R.F., Ablemann, C.H., & Fuhrman, S.H. (1996). The new accountability in state education reform: From process to performance. In H.F. Ladd (Ed.), *Holding Schools Accountable: Performance-based Reform in Education*, 65-98. Washington, DC: Brookings Institution.
- Fuhrman, S.H., & Elmore, R.F. (2004). Introduction. In S.H. Fuhrman and R.F. Elmore (Eds.), *Redesigning Accountability Systems for Education*, 3-14. New York: Teachers College Press.
- Garcia, G.E., & Pearson, P.D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of Research in Education*, 20, 337-391. Washington DC: American Educational Research Association.
- Goertz, M.E., Nabors Oláh, L., & Riggan, M. (2009). Can Interim Assessments Be Used for Instructional Change? (CPRE policy Brief RB-51). Philadelphia, PA: Consortium for Policy Research in Education.
- Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education*, 27, 25-68.
- Hanushek, E. A., & Raymond, M. E. (2004). *Does school accountability lead to improved student performance?* National Bureau of Economic Research Working Paper No. W10591. Cambridge, MA: Author.
- Herman, J. L. (2004). The effects of testing on instruction. In S.H. Fuhrman and R.F. Elmore (Eds.), *Redesigning Accountability Systems for Education*, 141-166. New York: Teachers College Press.
- Jovanovic, J., Solano-Flores, G., & Shavelson, R. J. (1994). Performance-based assessments: Will gender differences in science be eliminated? *Education and Urban Society*, 26(4), 352-366.
- Kober, N., Chudowsky, N., & Chudowsky, V. (2008). *Has Student Achievement Increased Since 2002? State Test Score Trends through 2006-07*. Washington, DC: Center on Education Policy.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5-16.
- Lane, S., Ventrice, J., Cerillo, T.L., Parke, C.S., & Stone, C.A. (1999). *Impact of the Maryland School Performance Assessment Program (MSPAP): Evidence from the principal, teacher, and student questionnaires (reading, writing, and science)*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.
- Renter, D.S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Jofthus, S., & Zabala, D. (2006). *From the Capital to the Classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.
- Stecher, B., & Barron, S. (1999). *Test based accountability: The perverse consequences of milestone testing*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Stecher, B.M., & Klein, S.P. (1997). The cost of science performance assessments in large-scale testing programs. *Educational Evaluation and Policy Analysis*, 19(1), 1-14.
- Supovitz, J.A. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10(2), 211-227.
- Supovitz, J.A., & Brennan, R.T. (1997). Mirror, mirror on the wall, Which is the fairest test of all? An examination of the equitability of portfolio assessment relative to standardized tests. *Harvard Educational Review*, 67(3), 472-506.
- Supovitz, J.A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education.

Research Notes

Penn GSE faculty and researchers explore the issues at the forefront of American education today, engaging in high-impact research, innovation, and training in public education, as well as in literacy, psychology, social policy, and higher and adult education. The following pages present a sampling of recent studies and findings from Penn GSE faculty and researchers.

What's Behind Being Behind?

In a recent study, Heather Rouse and John Fantuzzo examined the impact of five particular risk factors (poverty, child maltreatment, homelessness, low maternal education, and biological birth risks) on the educational outcomes of second-graders in the Philadelphia public school system.

To carry out their research, they used the KIDS data system, which, by integrating individual-level data across municipal agencies, enables researchers to trace children's entire histories of involvement with public service in the city.

Their multivariate examination revealed the presence of very high levels of each targeted risk—with poverty and birth risks most prevalent—and also analyzed the differing impacts of the *type* and *amount* of risks on both academic and behavioral outcomes.

The findings show that, first, *all five risk factors have a significant impact* on academic outcomes (i.e., reading and math proficiency and second-grade retention).

Second, *the type of risk matters*. For example, child maltreatment showed the highest level of risk across the greatest number of outcomes, and homelessness was one of the strongest indicators for high absenteeism, school suspensions, and teacher-rated low social skills. In addition, although the literature points to the negative impact of poverty, this study found that certain co-existing risks (maltreatment and maternal education) are a better predictor of academic outcomes than poverty per se.

Third, *the amount of risk matters*. Whatever the type, each additional risk experienced in early childhood increased the likelihood of poor academic and behavioral outcomes—most distinctly in the area of reading proficiency.

For Rouse and Fantuzzo, their findings carry significant policy implications. Put simply, what happens *outside the classroom* has a powerful impact on educational outcomes. “These findings underscore the injustice of holding public educators solely responsible for the educational well being of our most vulnerable children...” they write. “Educating children who are behind calls for the coordination of multiple service systems to meet these challenges.”

“Multiple Risks and Educational Well Being: A Population-based Investigation of Threats to Early School Success” appears in *Early Childhood Research Quarterly*, 24.

Testing the Cognitive Growth of Preschoolers

Cognitive growth during the preschool years is expansive and rapid, but most tests on the market, written for older children, aren't sensitive enough to measure learning over small intervals of time.

As part of a Head Start curriculum development project, Penn GSE researchers created just such a test. Called the Learning Express (LE), it was developed in field trials with over 3,400 Head Start children ranging in age from three to five-and-a-half years of age. Trained examiners administered LE in individual sessions lasting no longer than 30 minutes.

LE is an adaptive, skills-based test with four content domains—alphabet knowledge, vocabulary, listening comprehension, and mathematics—that include diverse subskills, aligned with the Head Start national standards.

Test items are ordered in ascending level of difficulty, starting with one that most children answer correctly. If the child does well on the first item, the test proceeds up the scale. After a specific number of incorrect responses, the assessor stops the session, thus cutting down on testing time.

To determine children's progress over time, reassessments are conducted approximately every two months. Individual growth-curve modeling demonstrated that the scores provide a highly sensitive measure of children's growth, controlled for age, sex, prior schooling, and language and special-needs status. In addition, multilevel modeling found that nearly all the variation in scoring was associated with the child's performance, not the examiner's.

"Measuring Preschool Cognitive Growth While It's Still Happening: The Learning Express," by Paul McDermott, John Fantuzzo, Clare Waterman, Lauren Angelo, Heather Warley, Vivian Gadsden, and Xiuyan Zhang, appears in *Journal of School Psychology*, 47.

Understanding Violence

As part of a community mental health outreach initiative, a prevention program called Understanding Violence was introduced into a Boston-area elementary school where students had been exposed to high rates of violence in their community. The program draws on students' educational and personal skills to practice positive alternatives to violence.

In their assessment of the program, Caroline Watts and colleagues surveyed 123 participating fifth-graders and found high levels of satisfaction with and learning from the program. In particular, the students responded positively to the program's use of diverse components and community engagement.

"*Understanding Violence: A School Initiative for Violence Prevention*," by Christina E. Nikitopoulos, Jessica S. Waters, Erin Collins, and Caroline L. Watts, appears in *Journal of Prevention & Intervention in the Community*, 37(4).

Is School Choice Rational?

The No Child Left Behind legislation guarantees parents the legal right to move their child from a “failing” school. Why then, asks Sigal Ben-Porath, do so few exercise that right?

In the debate over school choice, libertarians aim to bring the benefits of the marketplace into the educational system, while egalitarians hope for social justice. Whatever their end goal, both agree on the primacy of individual choice over regulation and intervention.

In Ben-Porath’s view, both ideas suffer a fatal flaw: the assumption of a rational actor. Drawing on insights from behavioral economics and ethnographic network studies, she sees choice as a far more complicated matter than simply weighing options and selecting the one that most closely aligns with one’s goals. Rather, she concurs with Nobel laureate Daniel Kahneman: “rational models are psychologically unrealistic.”

Empirical research on choice points to an array of influences that go largely ignored in the current debate, she continues. Notable among them are the way choices are presented, individual attitudes about risk and gain, and marginal channel factors that nonetheless can exert a strong influence on decision-making.

Consider the parents of the child whose school hasn’t made Annual Yearly Progress. Lacking adequate information about their options—and not being able to afford the cost of acquiring it—they will most likely make an intuitive decision based on immediate judgments (“My child feels fine there”).

But even those who do choose another school don’t base their decision on a classic cost-benefit analysis. Rather, parents rely on social networks, searching out anecdotal information from friends and relatives.

For policymakers, Ben-Porath’s analysis points to two areas of concern: the flow of information about available choices and the design of the choice set itself.

While Ben-Porath observes that the former is the “most amenable to policymaking intervention,” she adds that “it is not sufficient to educate individuals to choose wisely and inform them of their options. Society should be required to develop an equitable choice set.” To that end, she argues that choice policies need to take into account the family preferences as well as the societal aims.

“Choice policies need to avoid defaults,” she concludes, which effectively consign the children most in need to the weakest schools. Instead, policymakers should “develop a choice structure by which all must choose a school, and are properly supported in the process.”

“School Choice as a Bounded Ideal” appears in *The Journal of the Philosophy of Education*, 43(4).

The Best PR Firm You Never Heard Of

In the years after *Brown v. Board*, when many believed that historically black colleges and universities (HBCUs) were on the road to oblivion, one consultancy—the Oram Group—was working on major fundraising campaigns for these institutions.

Although virtually unexamined in the literature, the Oram Group can serve “as a lens through which we can view the development of modern day fundraising at black colleges,” observe Marybeth Gasman and Noah Drezner.

A for-profit firm that served progressive causes, the Oram Group was well ahead of its time in its hiring practices—by 1964, both women and blacks held high-level positions—as well as in its work with HBCU clients like Tougaloo College, Howard University, Atlanta University (now Clark-Atlanta), and Dillard University.

Unlike their competitors, the Oram Group urged black colleges to focus on alumni and the emerging black middle class and worked closely with leadership to build campaigns that would have a long-term impact. Finally, the Oram consultants encouraged their clients to promote the idea that, rather than being vestiges of segregation, HBCUs were vital to the whole of American society.

“A Maverick in the Field: The Oram Group and Fundraising in the Black College Community during the 1970s” appears in *History of Education Quarterly*, 49(4).

How Policy Effects Financial Aid

Public colleges and universities are faced with the challenge of balancing two competing goals: meeting student need and fostering academic excellence. Using data from the National Postsecondary Student Aid Award Survey, a recent study examined the impact of state policy on the distribution of financial aid awards at public institutions of higher learning.

Findings suggest that institutional behavior is increasingly mirroring state policy. That is, institutions in states with large need-based programs are more likely to grant aid on the basis of need than is the norm. Likewise, in states that deemphasize need, aid is more positively associated with merit.

For policymakers, these findings argue for careful attention to possible second-order effects of legislation. Thus, when state leaders enact policies that favor either merit or need, they may, in effect, be committing to those same policies at the institutional level. For institutional leaders, the results speak to “the importance of purposeful responses to state contexts.”

“Does Institutional Aid Compensate for or Comply with State Policy?” by William Doyle, Jennifer Delaney, and Blake Alan Naughton, appears in *Research in Higher Education*, 50(5).

Assessing Language Learning

For elementary-school children in South Korea, English is a compulsory subject. Moreover, the national curriculum emphasizes the importance of communicative competence—in particular, speaking—and, as a result, has begun to rely more on teacher-based assessments than on standardized testing.

To investigate how teachers observe and assess their students, what kind of criteria they apply, and how assessments may vary among teachers, Yuko Goto Butler recruited 49 English teachers—26 elementary and 23 secondary—who were participating in an in-service training program. After watching videotapes of students' group activities, the teachers were asked to assess their communicative performance, to identify the criteria they used in their evaluations, and to participate in a small-group discussion of their process.

In looking at the teachers' evaluations, Butler found substantial variations, both within and across groups. As for the question of evaluation criteria, teachers' reports of what traits they looked at were fairly consistent, but what emerged in the group discussion seemed to belie the notion of consensus. The elementary school teachers tended to avoid criteria-setting, relying instead on holistic judgments of students' performance. The secondary teachers, on the other hand, employed set criteria for their observations.

The differences between these two groups appear, Butler writes, “deeply rooted in their respective teaching contexts.” She

argues that the two sets of teachers should negotiate criteria that will ease student transition. Acknowledging that these two groups are customarily isolated from one another, she advises that policymakers should “create more opportunities for teachers to receive training together with teachers and different school levels and/or to exchange ideas and information among themselves.”

“How Do Teachers Observe and Evaluate Elementary School Students' Foreign Language Performance? A Case Study from South Korea” appears in *TESOL Quarterly*, 43(3).

Also from Yuko Butler

In a synthesis of information currently available in three Asian countries (South Korea, Taiwan, and Japan), Butler determined that previous research into the assessment practices of EFLES (English as a Foreign Language at Elementary Schools) programs has consisted of surveys and interviews focusing on the experiences and perceptions of students, teachers, and parents; studies using linguistic measures to compare the performance of EFLES students and non-EFLES students; and evaluations of students' language attainment based on criteria-based reference measures.

“Issues in the Assessment and Evaluation of English Language Education at the Elementary School Level: Implications for Policies in South Korea, Taiwan, and Japan” appears in *Journal of Asia TEFL*, 6(2).

History, Racial Justice and Higher Ed

Drawing on Critical Race Theory (CRT) as a conceptual framework, Shaun Harper and colleagues have developed a comprehensive examination of the policies created to ensure racial equity for African Americans in higher education and the development of those policies over time.

That history can be characterized as an “up-and-down’ struggle.” The authors provide an account of that struggle—from the establishment of Historically Black Colleges and Universities (HBCUs) for the education of freed slaves in the 19th century, to the desegregation initiatives sparked by the Civil Rights movement and the *Brown v. Board* decision, to the subsequent backlash and attempts to dismantle affirmative action. Gains notwithstanding, the authors conclude that “to characterize the current status of African Americans as inequitable would be a gross understatement.”

From an abundance of issues that have contributed to the demise of progressive policies, they home in on two: the pressures imposed on HBCUs and the policy conflicts confronting Predominantly White Institutions (PWIs). HBCUs, on the one hand, face declining enrollments as more of their traditional student base elects to attend PWIs; on the other, funding inequities have made it harder for HBCUs to compete for students, even as a series of court rulings has placed pressure on them to recruit more non-African Americans. Meanwhile, efforts to dismantle affirmative action at postsecondary institutions continue, with the most recent examples being two cases at the University of Michigan—*Gratz v. Bollinger* and *Grutter v. Bollinger*.

Understanding race as a social construct, CRT challenges the ways in which racial ideologies and various manifestations of white supremacy shape American life. Seen through its lens, the shifting landscape of policies regarding African Americans and higher education reveals the historical centrality of race to the present situation and promises a more nuanced understanding of current dilemmas.

Taking CRT’s “racial realism” approach, the authors argue that, by accepting that racism is endemic to the culture, African Americans can refocus their energies from unattainable goals to realistic strategies for addressing racial inequities on campus. They also outline the way in which interest-convergence—another of CRT’s central tenets—resonates through the history of African Americans in higher education. In their analysis of four areas—white missionaries and the founding of HBCUs; legislation like *Brown v. Board* and Title VI, state support of HBCUs, and affirmative action and college admissions—they ask, in effect, what motivated white support of these various initiatives.

If the racial justice promised in early policies is to be realized, policymakers and their institutional counterparts must understand the structural barriers that produce the current disparities in access and attainment. “While an elite group of African Americans have realized the promises of *Brown v. Board*,” the authors write, “the kinds of gains—social, political, and economic—that this group has made need to become more widely achievable to ensure access and equity. Much remains to be done.”

“Access and Equity for African American Students in Higher Education: A Critical Race Historical Analysis of Policy Efforts,” by Shaun Harper, Lori Patton, and Ontario Wooden, appears in *The Journal of Higher Education*, 80(4).

Also from Shaun Harper

Again drawing on one of the central tenets of Critical Race Theory—interest convergence—Shaun Harper considers how community colleges (and their athletic departments) would benefit from increasing the transfer rate of black male students to four-year institutions. “Race, Interest Convergence, and Transfer Outcomes for Black Male Student Athletes” appears in *New Directions for Community Colleges*, 147.

Data collected from face-to-face interviews with 143 black male undergraduates at 30 PWIs provide a counter-narrative to the dominant view of this population. Harper’s findings demonstrate that there is a significant overlooked population of high-achieving black male students on these campuses; that their simultaneous experiences of racism and success require them to adopt multifaceted coping strategies; and that to resist what Harper describes as “niggering,” they employ positive self-representation and immediate confrontation of racist stereotyping. “Niggers No More: A Critical Race Counternarrative on Black Male Student Achievement at Predominantly White Colleges and Universities” appears in *International Journal of Qualitative Studies in Education*, 22(6).

Racial minority students are less likely to engage in the high-impact educational experiences that are known to enhance student learning and improve degree attainment rates. In proposing the adoption of race-conscious engagement practices, Harper argues that institutions must actively set out to engage racial minority students and do so in ways that address their particular context and needs. “Race-Conscious Student Engagement Practices and the Equitable Distribution of Enriching Educational Experiences” appears in *Liberal Education*, Fall 2009.

North of the Border

The past 15 years have seen an increased rate of Mexican immigration into the United States, much of it to communities with few Latino residents. In an ongoing, four-year ethnographic study of one of these towns, Stanton Wortham and Penn GSE colleagues focused on residents' and immigrants' perceptions of this New Latino Diaspora.

The study site, a suburb of 30,000 in the Northeastern U.S. called Marshall, has seen its Mexican population grow from about 100 in 1990 to at least 6,000 in 2008. Twenty percent of the students in its school system are now Mexican.

In Marshall, as in other New Latino Diaspora towns, “more positive models of immigrant identity often have space to take hold,” the authors write. Thus, these new immigrants to Marshall do not necessarily experience the negative stereotyping typical of areas with a long-standing Latino presence.

Rather, Marshall residents tend to draw on models associated with earlier immigrant groups and describe Mexicans as hardworking and uncomplaining—a positive civic influence. Like the Italian and Asian immigrants before them—and in contrast to the local African-American population—these new immigrants are defined as “model” minorities.

This characterization does not hold, however, when the topic turns to education. Overall, Marshall's educators had fairly low expectations for their Mexican students, particularly in comparison to Asian Americans, who were described as being “big on education.”

“Educators in New Latino Diaspora towns face a challenge,” the authors write. Often under-resourced, schools in New Latino Diaspora towns struggle to serve a population of students with which they have limited experience; most likely, they do so in a context where those students are not expected to excel academically.

Those struggles notwithstanding, the authors conclude by posing the real challenge to educators: “Given the flexibility of social identification in the New Latino Diaspora, might educators have opportunities to identify Mexican students not only as hardworking but also as academically promising?”

“Mexicans as Model Minorities in the New Latino Diaspora,” by Stanton Wortham, Katherine Mortimer, and Elaine Allard, appears in *Anthropology & Education Quarterly*, 40(4).

It Takes a School To Raise a Community

A recent article tracing the emergence of the community school movement in the United States points to East Harlem's Benjamin Franklin High School in the 1930s and 1940s as exemplary. Its principal, Leonard Covello, "emphasized the school as a means for social problem solving and for training students in effective democratic citizenship." Lacking sustained support, however, Covello's vision could not survive his tenure.

Recent decades have brought promising signs for a reinvigorated community school movement—including nonprofit-school collaborations and university-community partnerships. By the mid-2000s, a number of jurisdictions had begun sponsoring community schools that provide health, family-support, and youth-development services.

"In each case," the authors write, "a nonprofit played a lead role—removing the burden from the schools of developing partnerships, securing funding, and coordinating services."

"The Enduring Appeal of Community Schools: Education Has Always Been a Community Endeavor," by Lee Benson, Ira Harkavy, Michael Johaneck, and John Puckett, appears in *American Educator*, Summer 2009.

Academic Alliances That Work

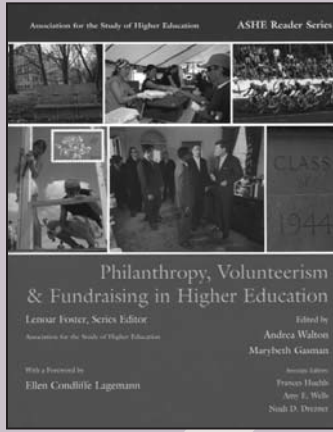
How can colleges and universities respond to emerging societal needs in the current economic downturn? Peter Eckel and Matt Hartley offer one answer to that question: strategic interorganizational partnerships. These "curricular joint ventures" are collaborations that create an academic venture beyond the capacity of any one partner alone. Eckel and Hartley were interested in how institutions develop rules to reconcile disparate organizational cultures.

Applying a comparative case study methodology, they examined three partnerships: a dyad (Virginia Tech-Wake Forest University School of Biomedical Engineering and Sciences), a medium network of five institutions (OneMBA), and a large ten-institution network (Great Plains IDEA).

Their findings suggest that successful partnerships are characterized by a willingness to recognize and address inherent tensions between individual and collective needs; a commitment to ongoing face-to-face interaction; and the creation of a shared identity.

Contrary to the advice offered by the management literature, which typically focuses on operational concerns, this study found that the key for success in curricular alliances is "forging relationships and forming a shared identity."

"Developing Academic Strategic Alliances: Reconciling Multiple Institutional Cultures, Policies, and Practices" appears in *The Journal of Higher Education*, 79(6).

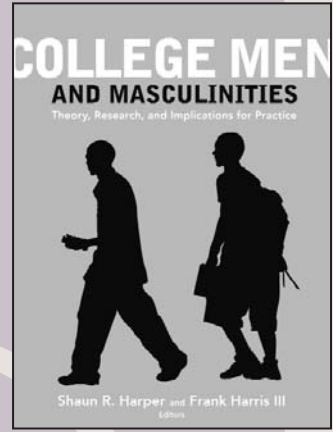


Andrea Walton and Marybeth Gasman (Eds.) with Fran Huehls, Amy Wells, and Noah Drezner
Philanthropy, Volunteerism & Fundraising in Higher Education

ASHE Readers Series, 2009

The reliance of academic institutions on private funding has been a distinctive feature of higher education in the United States dating back to colonial times. Presenting original introductory essays, research articles covering major approaches and topics, and primary documents for teaching, this volume addresses key issues in educational philanthropy and higher education.

Winner of the John Grenzebach Research Award for Outstanding Published Scholarship, awarded by the Council for Advancement and Support of Education



Shaun R. Harper and Frank Harris III (Eds.)
College Men and Masculinities: Theory, Research and Implications for Practice

Jossey-Bass, 2010

This comprehensive handbook synthesizes research and literature on college men, discusses implications for practice, and addresses larger topics on race, class, and gender in higher education. Written for faculty, staff, and administrators, this book provides analysis and practical prescriptions for reversing problematic trends and outcomes among male undergraduates.

Ira Harkavy and Matthew Hartley (Eds.)
Universities in Partnership: Strategies for Education, Youth Development, and Community Renewal

Jossey-Bass, 2009

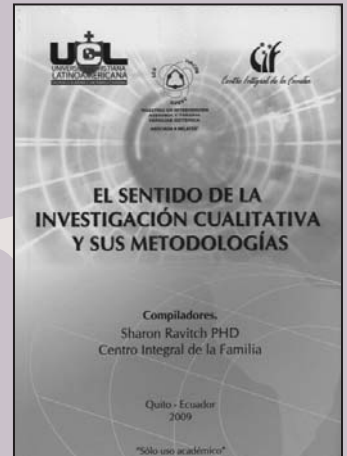
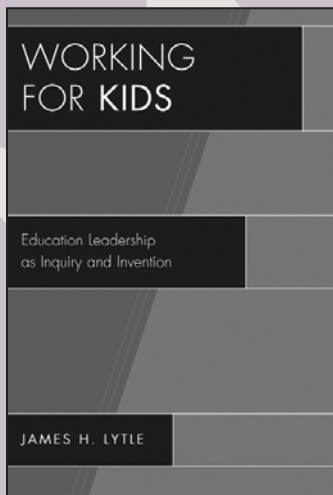
While the civic engagement movement has spawned a number of intriguing university-community partnerships, these initiatives face many challenges: overcoming traditional ivory tower thinking, creating comprehensive approaches, and sustaining democratic collaborations. Examining case studies from five universities—SUNY Buffalo, Indiana University-Purdue University Indianapolis, the University of Dayton, Widener University, and Penn—this volume demonstrates that these partnerships can succeed.



James H. Lytle
Working for Kids: Education Leadership as Inquiry and Invention

Rowman & Littlefield, 2010

Drawing on the author's experience as a school administrator and academic, *Working for Kids* critiques the current strategies for improving school leadership. Lytle argues that, rather than emphasizing evaluation and standards determined by policymakers and professors, these initiatives should address questions of performance-driven leadership. How does one learn to lead? How does one become a leader? How does one teach others to lead? What does it mean to lead for learning?



Sharon Ravitch (Ed.)
El Sentido de la Investigación Cualitativa y sus Metodologías (The Meaning of Qualitative Research and Methods)

Centro Integral de la Familia, 2009

This compilation includes Spanish-language articles that outline the uses of and approaches to qualitative research broadly and also specifically address issues for systemic family therapists working with rural and urban families from indigenous and non-indigenous backgrounds. Published in Quito, this book is intended to introduce qualitative research for practitioners and students in Ecuador.



Graduate School of Education
PennGSE

3700 Walnut Street
Philadelphia, PA 19104-6216

Non-profit Org.
U.S. Postage
PAID
Permit No. 2563
Philadelphia, PA

THE PENN GSE PODCAST



Perspectives on education and
social science from Penn GSE

WWW.GSE.UPENN.EDU/PODCAST