

educational HORIZONS®

was founded in the spirit of academic excellence in order to provide leadership in addressing educational, social, and cultural issues of national and international significance, and to enhance the status of educators by providing a recognized forum for sharing new perspectives, research findings, and scholarly essays.

CONTRIBUTING EDITORS

Gary Clabaugh
La Salle University

Edward G. Rozycki
Widener University

Martha M. McCarthy
Indiana University

Wade Carpenter
Berry College

Audrey Ricker
University of Arizona

GRAPHIC DESIGN
Raven Bongiani

COPYEDITING
Tony Brewer, The Same Page
Stephen S. Gottlieb

ILLUSTRATION
Barbara Murray

PI LAMBDA THETA
Rita J. Jones, President
P. O. Box 6626
Bloomington, IN 47407-6626
Phone (812) 339-3411
Toll-Free (800) 487-3411
Fax (812) 339-3462
office@pilambda.org
www.pilambda.org

Departments

THE CUTTING EDGE

A Most Essential Question: Who Is Truly Educable? 2
by Gary K. Clabaugh

BEHIND EVERY SILVER LINING

The Other Side of No Child Left Behind 7
by Wade A. Carpenter

ON BALANCE

Is Banning Holidays the Only Way? 12
by Cory Gann

Features

TESTING CONCERNS

Characteristics of an Effective Student Testing System 19
by Richard P. Phelps

Perils of Standardized Achievement Testing 30
by Thomas M. Haladyna

Testing for Justice 44
by Edward G. Rozycki

DATA-BASED DECISION-MAKING

Data-based Decision-making: Three State-level Educational Leadership Initiatives 57
by Van E. Cooley, Jianping Shen, Deborah S. Miller, Peter N. Winograd, John Mark Rainey, Wenhui Yuan, and Lisa Ryan

Challenges in Data-based Decision-making: Voices from Principals 65
by Patricia L. Reeves and Walter L. Burt

THE CUTTING EDGE

A Most Essential Question: Who Is Truly Educable?

by Gary K. Clabaugh

What is the most essential question we can ask about schooling? Try this one: how many people are truly educable?

Reason and Understanding

What does it mean to say that someone is “educable”? Special education provides a frame of reference. Traditionally, mentally handicapped people are classified as “educable mentally retarded.” The more severely impaired are grouped as “trainable mentally retarded.”

What’s the difference between being “educable” and “trainable”? That is a complicated issue, but for the purposes of this consideration let’s say that an “educable” person must be “capable of being improved in ways that depend on reason and understanding.”¹ A person who is “trainable” is largely incapable of being improved in that way. Of course, educable individuals can be trained, but not vice versa.

Faith in Education

Americans have long had a peculiar faith in the problem-solving power of education in general and schooling in particular. They are widely regarded as the answer for nearly every difficulty. Consequently, our public schools are expected to solve a full spectrum of such difficulties as the cultural integration of immigrants, the eradication of poverty, the

educational HORIZONS® (ISSN 0013-175X) is published quarterly by Pi Lambda Theta, Inc., P.O. Box 6626, Bloomington, IN 47407-6626. Pi Lambda Theta membership, which includes a subscription to educational HORIZONS®, is \$40 per year. Nonmember subscriptions are available for \$18 per year; U.S.; \$25 per year; Canada and international. Periodicals postage paid at Bloomington, Indiana, and other mailing offices. Single copies: U.S. \$5, Canada \$6.50, and international \$8, plus \$1.50 postage. POSTMASTER: Address changes should be mailed to: Pi Lambda Theta, P.O. Box 6626, Bloomington, IN 47407-6626. All claims must be made within four months of publication. Back volumes available on microfilm from National Archive Publishing Co., 300 Zeeb Rd., P.O. Box 998, Ann Arbor, MI 48106-0998; (800) 420-6272; outside N.A., (734) 302-6500; www.napubco.com. Indexed in Current Index to Journals in Education (ERIC). Selected back issues available online at www.pilambda.org. Copyright 2006, all rights reserved, by Pi Lambda Theta. Opinions expressed herein are those of the authors and do not necessarily reflect the official views of Pi Lambda Theta. educational HORIZONS® is a trademark of Pi Lambda Theta.

elimination of racial injustice, the decline in national competitiveness, and so forth.² Indeed, the variety of problems thought susceptible to educational solution is nearly as inexhaustible as human troubles.

In Philadelphia, for instance, residents of the City of Brotherly Love have been murdering one another with surprising enthusiasm this year. In fact, it looks as if Philadelphia is heading for a record four hundred homicides and as many as five thousand shootings in 2006. Just the other day, four individuals were gunned down at a subway stop and an adolescent was shot to death for making fun of another teen's sneakers.

Americans look at such seemingly senseless violence, wring their hands, and ask, "What is to be done?" And in the end, after much discussion and many failed "solutions," most conclude that the only way to stop the madness is, you guessed it, education.

Take the Philadelphia sneaker homicide, for example. No sooner had the blood been washed from the sidewalk than an earnest citizen e-mailed the editor of a local newspaper to suggest that more conflict-resolution classes are the solution. Teach them how to resolve conflicts peacefully, and they will quit killing one another, was the message. That's the way of it in America. Education is the answer for whatever difficulty society faces. But is it really?

The Triumph of Hope over Experience

Certainly much human misery would be prevented if people could be taught to think more deeply and effectively. But is the common failure to do so a consequence of ignorance and lack of education, as many suppose? Perhaps it is a widespread lack of capacity instead.

For education to be a cure, much less the cure, the majority of humans must be capable of sufficient reason and understanding to be improved by that means. Suppose, just suppose, that is not the case. Perhaps optimism concerning education's potential represents the triumph of hope over experience. Possibly a great many humans are not truly educable, but merely trainable.

That statement sounds heretical to those accustomed to the obligatory and oftentimes pathological optimism that surrounds schooling. But there is evidence to support such a view. For instance, the Harris Poll recently reported that despite repeated reports that no weapons of mass destruction have been found in Iraq, the belief that the country possessed such weapons has substantially increased since last year.³ That's right: despite massive evidence to the contrary, the number of Americans who think that such weapons were there has actually gone up. As a matter of fact, in February 2005 only 36 percent thought Iraq had been so armed before the war; but by July 2006 fully 50 percent believed it. Does that sound like a conviction that grew out of widespread public capacity for

reason and understanding—or even for paying attention? On the other hand it might be that those who changed their minds about those weapons of mass destruction desire to rationalize their own original enthusiasm for the war and to justify the tremendous costs it has generated. In short, this seeming evidence of public credulity might just be people being human, all too human.

We can also profitably consider the success of political campaign strategies based on the supposition that no one ever lost an election by underestimating the intelligence of the American people. (That's a derivation of H. L. Mencken's observation: "No one ever went broke underestimating the intelligence of the American people.") In Pennsylvania, for example, a \$3.5 million TV ad blitz by Sen. Rick Santorum recently cut the substantial lead of his challenger, Bob Casey, to just six percentage points. Its chief feature is referring to Casey as "Bobby" to make him seem juvenile and inconsequential. Doubtless Casey aides are trying to come up with a similar strategy. Why? Because ads like this work.

Does this suggest there is a great deal of deep thought going on out there? Political propagandists have learned to play on emotions such as fear of the unknown, the alien, and the complex. Moreover, the simplicity they offer is beguilingly attractive to a public that has to reach conclusions based on imperfect information and deliberate disinformation.

A widespread lack of public discernment is not confined to politics and foreign affairs. Consider P. T. Barnum's observation: "There's a sucker born every minute." Now consider television infomercials. Aren't they often the electronic equivalent of old-time snake-oil sales? Omega-3 fish oil, we are assured, can cure an amazing range of maladies. And purging one's bowels with Clean Sweep evidently has the same beneficial effects on the human body that emptying a full sweeper bag can effect for a clogged-up Electrolux. Consider also the televangelists who hawk "miracle spring water" and prayer cloths with apparent miraculous powers or the infomercial "pastors" who successfully assure viewers that generous giving will eliminate their financial troubles. One particularly persuasive televangelist, for instance, lives in a multimillion-dollar California beach home and owns a private jet—a lifestyle comically remote from that of Jesus.

Multiple millions of dollars are spent buying TV time to peddle such nostrums, and many more millions are realized in consequence. When considering the general educability of Americans, what shall we make of that sort of thing?

For reasons of space, supplying additional instances of a widespread lack of public discernment must be left to the reader. Suffice it to say that any list of human follies is likely to be a long one. Does that evidence a widespread lack of educable people?

Stupidity Is Not the Only Impediment to Reason

Stupidity is not the only thing that prevents intelligent reflection. Individuals often have considerable native intelligence and still cannot or will not think because they are too emotionally needy. Such persons are not too dumb to think straight; they are too angry, scared, insecure, guilty, depressed, and so forth.

Also, many intelligent but emotionally needy people willfully shut off their intelligence to gain emotional reassurance from one true belief or another. The people who joined Jim Jones's congregation, the Branch Davidians, or the Heaven's Gate cult, for example, were not necessarily stupid. Often it was their emotional needs that got the better of them and caused them to stop thinking. Similarly, our prisons are overflowing with natively intelligent people who, for various reasons including childhood neglect and abuse, simply cannot control their own behavior even when it clearly is in their own best interests to do so.

Culture can prove another barrier to reason. Some cultures greatly facilitate intelligence by providing rich resources for reflection, but others do not. Consider what might happen to one's capacity for reflective thought if one were raised in a fundamentalist state such as Saudi Arabia, for instance, where Wahabism, an intolerant Sunni fundamentalist religious movement, largely dominates the culture. It is not hard to imagine that high native intelligence might be channeled into ingeniously destructive fanaticism of the type evidenced by Osama bin Laden, himself a Saudi.

Schooling, Conformity, and Technical Skills

We also should not assume, as many do, that increased schooling necessarily equals greater capacity for reasoning and understanding. Schooling is often less about critical reasoning than it is about conformity and mastering technical skills. Consider the abundance of "scientists" who eagerly apply their technical competence to creating unimaginably vicious weapons of mass destruction. Are the men or women who apply their knowledge of biology to perfect a vaccine-resistant plague virus really reasoning the thing through as well as they should?

On a more mundane level, what evidence is there that the average M.B.A., or certified teacher for that matter, is significantly more reasonable or thoughtful than those less schooled? Those with an educational imprimatur have hopefully mastered certain techniques and learned to mouth the right platitudes, but can they think more deeply and well?

Consider Lyndon B. Johnson's "best and brightest": Secretary of Defense Robert McNamara, B.A. UC Berkeley, M.B.A. Harvard; Special Assistant to the President McGeorge Bundy, Groton, Yale, Harvard; and Secretary of State Dean Rusk, Rhodes scholar, Oxford, UC Berkeley. What

did those highly educated can-do guys accomplish? They bogged us down in a losing ten-thousand-day war in Vietnam. Forty-eight thousand Americans died; another 304,000 were wounded. Many more were maimed in spirit. In fact an estimated nine thousand suicides were a direct result of the war.⁴ There were an appalling 5.1 million Vietnamese casualties.⁵ One hundred eleven billion dollars was wasted and America was torn apart domestically. Does that sound like the work of men with superior reasoning and understanding?

Then there is George W. Bush. He has a B.A. in history from Yale University and an M.B.A. from Harvard. What evidence is there that his reasoning and understanding are superior in consequence of that experience?

In sum, schooling is not necessarily the friend of deep and effective thinking, even when it is accomplished in the most hoity-toity environs. There are other variables.

What Do You Think?

How many people are truly educable? Is the widespread American belief in education misbegotten or essentially realistic? Does the general population have the intelligence and the emotional capacity to be educated in the sense we've used it here, or was Mencken right when he contemptuously labeled average Americans the "booboisie"? You decide.

Notes

1. John Hyman, "Is Beauty in the Eye of the Beholder?" *Think*, Spring 2002.
2. See Henry J. Perkinson, *The Imperfect Panacea: American Faith in Education*, 4th ed. (New York: McGraw Hill, 1995).
3. The Harris Poll #57, July 21, 2006.
4. Centers for Disease Control and Prevention, "The Vietnam Experience Mortality Assessment."
5. Senate Committee on Veterans' Affairs, *Oversight on Post-Traumatic Stress Disorder*; 100th Cong., 2nd sess., 1988, 17, quoted in "How Many People Died in the Vietnam War?" <<http://answers.google.com/answers/threadview?id=5096>>.

Gary K. Clabaugh, Ed.D., is a professor of education at La Salle University.

The Other Side of No Child Left Behind

by Wade A. Carpenter

The surest way to discredit the public school is to leave no child behind. The second surest way is to make the school a “safe and nonthreatening environment” for psychopaths and morons. The third surest is to try to teach the “whole child” in any single setting. The conservatives know this; that’s probably why they’re doing it. And in a fit of unconscious bipartisanship seldom equaled in our contentious nation’s history, the liberals have been working on the same project for decades, albeit more stupidly. The sad fact is that public schools just cannot be all things to all children, nor should they be a total program for producing “whole children.”

Please forgive my unaccustomed bluntness. I reckon I’d better explain that when I use the coarse word “morons,” I am not referring to those who cannot learn—they deserve every kindness. Rather, I mean those who *will* not learn and who are gratuitously disruptive. They need to be elsewhere, as do those who are cruel and without working consciences—the psychopaths. I believe in educating every child, but I no longer believe public schools and “regular” classrooms can accomplish that. Education is a right, but alas, schooling is the price we pay for waiving that right. It seems to me that compulsory education, as distinct from compulsory schooling, requires multiple alternatives and viable choice. Given the extraordinary diversity of purposes for schools and definitions of education in a democracy, a rich diversity of educational offerings makes sense, so I have no problem with private schools, and I’m favorable toward vouchers, with reservations.¹ However, decent people just don’t work toward choice or privatization in ways that would unnecessarily hurt kids and teachers, and No Child Left Behind does precisely that.² Hence, I don’t blame the Bush administration for its goal, but I do blame it for its method. On the other hand, liberal-progressive insistence on “whole child” teaching in a government-school setting—an idea that can be more totalitarian than charitable, and is notoriously difficult to square with the classically liberal philosophy of freedom—has only added an impossible ideal to a flawed institution.³ This leaves me blaming the liberals for their goals, and *not* their methods, which are generally ethical, informed, and skillful.

So here's the bad news: I fear that this column marks a turning point for me, in which I make a long-dreaded shift from "maverick" to "renegade." In my previous columns I have tried to take the high road, assuming the best intentions of whomever I was disagreeing with, and moderating my language to be constructive and charitable. (For instance, see my mixed review of John Taylor Gatto's *Underground History of American Education* in the winter 2005 issue of this journal. I saw the book as basically factual, but one-sided and angry. I believed then that Gatto was correct but wrong: that there was far more good going on in our schools than harm.⁴) Over the past year or so, my opinion has changed. I've encountered the most despicable miseducation I've seen or even heard of in thirty-three years of teaching—so bad, in fact, that I'm no longer willing to be tactful. I ended the year telling one of my student teachers and his cooperating teacher: "If I were a smart kid in this class, I'd either drop out or commit suicide." (To their credit, they agreed wholeheartedly.) My junior-level students, in a methods course that is entirely positive—simple how-to-do-it stuff, nothing to do with problems and policies—ended their field experiences this year with the gloomiest countenances ever. Student after student, they confessed to me that what they were seeing "just isn't worth doing," and from what they described, I couldn't argue with them.

About the only thing I could advise was that "Smart people don't make career choices based on March morale." True enough, but still pretty lame. What we were seeing was wooden, rote learning and brainless, boring teaching, worthy of the contempt of any free American. What we were seeing, almost uniformly, was test-driven minimalism, with the slower and resistant kids monopolizing the time of frustrated and surly teachers, and the brighter kids sitting quietly, bored stupid. In a few cases the teachers were clearly blameworthy, burnt-out, or useless to begin with, but mostly we saw decent people trying heroically to be blessed exceptions—heroically, but with depressingly little success against a system that demanded basic knowledge but penalized advanced thinking. And we saw an unprecedented teacher attrition rate, with teacher dropouts' places taken increasingly by untrained "drive-by" teachers. And that made my students' gloom that much worse: nobody looks forward to a life like that, nor do they want to impose such unkindness and mediocrity on children. I pray that my students and I just got two bad-luck sets of assignments, and that what we saw is not generalizable to either the local schools or the rest of the country. Unfortunately, I'm becoming increasingly alarmed that it may be.⁵

It should be no surprise, since the famed Christian writer C. S. Lewis identified the process almost fifty years ago. In "Screwtape Proposes a Toast" (a sequel to *The Screwtape Letters*), Lewis took the role of a senior

devil advising fledgling demons how to make people mediocre enough to be suckered into hell:

The basic principle of the new education is that dunces and idlers must not be made to feel inferior to intelligent and industrious pupils. That would be “undemocratic.” . . . At universities examinations must be framed so that nearly all students get good marks. . . . At schools, the children who are too stupid or lazy to learn languages and mathematics and elementary science can be set to doing the things that children used to do in their spare time. Let them, for example, make mud pies and call it modeling. But all the time there must be no faintest hint that they are inferior to the children who are at work. . . . Children who are fit to proceed to a higher class may be artificially kept back, because the others would get a *trauma*—Beelzebub, what a useful word! The bright pupil thus remains democratically fettered to his own age group throughout his school career, and a boy who would be capable of tackling Aeschylus or Dante sits listening to his coeval’s attempts to spell out A CAT SAT ON A MAT. . . . And the teachers . . . will be far too busy reassuring the dunces and patting them on the back to waste any time on real teaching. We [devils] shall no longer have to plan and toil to spread imperturbable conceit and incurable ignorance among men. The little vermin will do it for themselves.⁶

This is precisely what we are seeing, and looking at those bright children wasting away in those wretched classes infuriates me now just as seeing those minority kids wasting away infuriated me years ago. Time to declare war. The good news is I think it’s just possible to win this one. For now, let’s start with one idea and do whatever I can fit into the rest of my word allotment for this issue, then continue to some more radical and thoroughgoing ideas in coming issues.

What schools are says little about what schools could be and almost nothing about what schools should be. Schools can and should be safe places in which worthwhile learning is the principal goal, and becoming worthwhile adults is the principal outcome.⁷ So allow me to suggest that educators have the responsibility to protect kids from three intellectual vices: ignorance, stupidity, and silliness. Ignorance is when one doesn’t know much. Stupidity is when one knows only what one has been told. And silliness is when one knows only what one wants to know.⁸ Granted, no school is going to be 100 percent effective in protecting kids intellectually any more than we will be 100 percent effective physically. Just as the occasional nitwit will get a black eye despite the best adult supervision, so some nitwits will not achieve well in schools, ending up

still ignorant, stupid, and/or silly. While the “no child left behind” ideal is a moral improvement over the old “somebody’s got to dig the ditches” mentality, it is not a moral victory, nor is it likely to become one. It’s much more likely to generate lots more bad press (we call that “accountability”) for the public schools, which I suspect is its real purpose. And, frankly, I plan to do my part: I’m no longer going to tolerate what I’ve been seeing the past couple of years, nor will I be nice about it. Heaven knows I don’t expect perfection from individuals or justice from institutions, but like military veterans everywhere, I tend to give more credibility to the grunts on the line than to the REMFs (Rear-Echelon Master Foulups, or something like that), and to place lots more blame on their institutions and lay responsibility for fixing it on those mid-level managers wise enough to disdain further promotion.⁹ The measure: if a school can’t do at least as good a job at protecting kids—and teachers—intellectually as it does protecting them physically, it has no business being in business.

With this “standard” to meet, the schools will need to recruit and keep smart, skilled teachers who are experienced enough to be wise teachers also. As discussed in this column previously, the supposed “teacher shortage” is baloney.¹⁰ Although spot shortages surely exist (not enough science teachers here, not enough special ed teachers there, et cetera), the real problem is teacher retention, especially the retention of good teachers. So I propose a “friendly amendment” to No Child Left Behind’s “highly qualified teacher” provisions, one that I daresay no legislator could safely oppose but that all the REMFs will try to bury. Let’s add a provision requiring all public schools to be “highly qualified schools” by 2012, with one criterion being that on-site teacher longevity averages at least five years. Imagine how different schools would be if the pressure on administrators to seek, develop, and keep serious teachers (and quickly get rid of ignoramuses, incompetents, and fools) were as strong as the current pressure is to raise the test scores of thugs. Imagine how different schools would be if administrators were under that kind of pressure to turn their schools into places where teachers want to stay. Imagine that, and it becomes possible to imagine schools as places where students want to be, rather than have to be. I believe administrators would find ways to do that, and I know they’d have to start now.

Notes

1. See Wade A. Carpenter, “The Other Side of School Choice,” *Educational Horizons* 83(2) (2005): 87-91.

2. While I think the literature against No Child Left Behind is blemished by many self-serving rants by professional educators (including some of my own), there is a substantial literature of substantive critique. One of the more interesting treatments is Lowell C. Rose’s “No Child Left Behind: The Mathematics of Guaranteed Failure,” *Educational Horizons* 82(4) (2004): 121-130.

3. The literature on the liberal definition of freedom is voluminous. Basic references

must include John Stuart Mill's *On Liberty* (1869; New York: Bartleby.com, 1999); Ivan Illich's *Deschooling Society* (New York: Harper & Row, 1972); and John Dewey's *Democracy and Education* (New York: Macmillan, 1916). I also recommend Barry L. Bull, Royal T. Fruehling, and Virgie Chattergy's *The Ethics of Multicultural and Bilingual Education* (New York: Teachers College Press, 1992) for an education-oriented view; John Taylor Gatto's *Dumbing Us Down: The Hidden Curriculum of Compulsory Schooling* (Philadelphia: New Society, 1992) for a more radical educationist analysis; Kenneth L. Grasso, Gerard V. Bradley, and Robert P. Hunt's *Catholicism, Liberalism, and Communitarianism* (London: Rowman & Littlefield, 1995) for a religious perspective; John Kekes's *Against Liberalism* (Ithaca: Cornell University Press, 1997) for a philosophically technical attack; and Peter Augustine Lawler's "Technological Nihilism and Natural Law: Science, Morality, and the Law Today," *Educational Horizons* 83(4) (2005): 564–271, for a freewheeling anti-Deweyan diatribe.

4. Carpenter, "The Other Side of School Choice."

5. Especially alarming to me has been a series of articles by Ann Doss Helms from May 30 to June 2, 2006, in my hometown newspaper, the *Charlotte Observer*, detailing the miserable collapse of a once-ballyhooed school system. I was there in the middle of its so-called glory days in the 1970s and '80s, when it got national raves for "making busing work." While that glory was somewhat exaggerated (see Frye Gailliard's hagiographic *A Dream Long Deferred* [Chapel Hill: University of North Carolina Press, 1988]), it was many notches above the "bright flight" catastrophe that Judge Howard Manning, Jr., described from the bench as "academic genocide."

6. *The Screwtape Letters and Screwtape Proposes a Toast* (Westwood, N.J.: Barbour and Company, Inc., 1961), 180–181.

7. The fall 2004 issue of *Educational Horizons* has a number of good articles addressing physical and emotional safety in schools.

8. We used to call this by terms such as "narcissism," "solipsism," "hubris," and "egoism"; the founding fathers feared it as "factionalism"; and modern politicians pretend to deplore it as "partisanship." More recently, of course, we have called it "postmodernism" and "constructivism." By whatever name, it still amounts to "same foot, different pile," and it is just plain silliness.

9. George Becker's *Assassin's Gate: America in Iraq* (New York: Farrar, Strauss, and Giroux, 2005) puts it well: "In Iraq as in Vietnam, I continually found more insight among midlevel civilians and military than at the top—because the political pressure at that altitude is low enough for clear thinking to take place, and because their intellectual candor made professional advancement less likely" (p. 304). A couple of depressing examples of why I trust teachers' ground-level input over their better-informed bosses' were in yesterday's and today's local newspapers—and that's probably about as far as anybody else would have to look anywhere in the nation! The July 12 *Rome* [Ga.] *Tribune* gives this year's AYP results, and despite all the hype and PR of the past two years, the local schools are, to put it nicely, turning out cheap labor, and that's about it. Today's paper describes the sentencing of our state's previous superintendent to eight years for stealing \$600,000 from the Georgia School for the Deaf. The School for the Deaf, for God's sake!

10. Wade A. Carpenter, review of *Who Controls Teachers' Work?* by Richard Ingersoll, *Educational Horizons* 84(2) (2006): 69–77.

Wade A. Carpenter, Ph.D., is an associate professor of education at Berry College, Mount Berry, Georgia.

ON BALANCE

Is Banning Holidays the Only Way?

by Cory Gann

In teacher education, do we mindlessly assign lessons about holidays? What do we want to see happening with regard to holidays in schools? What parameters do we give students for addressing holidays? What are we after, anyhow?

The holiday question continues to be one of public school's incessant enigmas. Like most educational paradigms, holiday treatment has rotated through dialectical swings of attention and neglect, deliberate curriculum inclusion, as well as conscious curricular omission. Holiday volume has been turned down for December's special days and ratcheted up several decibels for Groundhog Day and the 100th Day of School. An intractable need to celebrate something is pitted against a growing awareness that continuing allegiance to a Hallmark holiday menu invariably shuts more and more families out of real connection to these events. As these matters scale up through the power structure, policy becomes a marginalizing factor emblemized by occasional school board manifestos to ban Halloween from elementary school classroom celebration lesson plans.

Public schools, especially at the elementary, and particularly at the primary level, need to leave holidays to teachers and to families and children. At the same time, teachers need to return to the important question that is supposed to drive curriculum policy—what are we trying to accomplish here above and beyond just getting through one more special day?

A Little History

Defying a monocultural pattern of school celebration, a wave of diversity in the 1960s and 1970s promulgated factoid information about and attention to special days rooted in minority traditions. Into the cultural amnesia came festivalized planning for Hanukkah, Chinese New Year's, Cinco de Mayo, and (when it became nationalized) Martin Luther King Jr.'s birthday. Only an unexpected thing happened: the inclusion-contribution approach was eventually exposed as devoid of consciousness or willingness to face genuine diversity issues such as racism,

cultural imperialism, or dominant-group centrality. Increasingly, holiday curriculum, along with food, dress, and festivals, became part of “tourist curriculum” (Derman-Sparks 1988). As Carol Brunson Phillips mused:

Will the study of culture and the appreciation of cultural differences enable our children and families to overcome their struggles for power over their own quality of life? . . . My belief is that they will not. My belief is that culture is not the problem, nor that differences are nor that diversity is a root cause of inequality. . . .

Rather than *diversity* itself, it is the ways in which the major institutions of this country have responded to culturally, racially, and ethnically diverse people that is the major source of our condition of social, political, and economic inequality. (Phillips 1989)

Examining institutional (especially school) responses to diversity sowed the seeds of transformative concepts such as cultural sensitivity, cultural relevance, and social reconstruction. Teachers were faced with new maxims regarding holiday pedagogy: Don’t make the holiday your main focus about a cultural group; be careful about the “exotic”; be sensitive to Jehovah’s Witnesses and others who don’t celebrate; and especially, include all the holidays or don’t include any. That last one touched a chord with harried educators and heralded the era of holiday-free curriculum.

De-emphasizing holidays, however, does little to recognize people’s visceral need to celebrate, a need arguably activated at a very young age like a Chomskyan language device or perhaps acquired early in social development as young children assimilate birthday songs and gifts. One reason that the holiday question continually resurfaces is children’s inherent disposition toward special days, despite efforts to standardize and commercialize the school day. A primary-age child’s birthday is prominent, and the communal veneration for marking passages and milestones is reinforced even by, for example, observing the 100th day of school. Children who have never heard of Dewali or Ramadan know all about Punxsutawney, Pennsylvania, and the meaning of the groundhog’s shadow. In this way, largely uncontroversial and safe holidays have squeezed themselves into curriculum plans, backed up by a set of entrenched projects, regurgitated year after year. One need only visit a Halloween or St. Patrick’s Day educational Web site to learn about myriad craft make’n’take activities to keep primary kids busy constructing one more artifact to go on the tree, or on the refrigerator, or in the memento box.

Although craft fabrications do provide children a break from standardized, scripted lessons, such activities also inherently pose problems. First, an unintended inauthenticity inevitably takes hold. All kinds of spiritual Native American artifacts—totem poles, sand paintings, kachina dolls—suffer trivialization when presented as crafts to be

copied. At the same time, the preferred art lesson or writing prompt often contributes to inaccurate conceptions by omission, if not commission. The Martin Luther King, Jr., “Man of Peace” story becomes historical untruth when presented outside the context of the civil rights movement, and when this great figure is reified as an autonomously heroic doer of deeds.

Perhaps just as important is the tunnel-visioned constraint that activity-box observances impose on the learning community. The automatically and unconsciously chosen art project can limit available options when a calendar event arrives. Busywork thing-making is the classroom’s first and often only affirmation of the holiday’s existence. Think of Thanksgiving, or Passover, or Dr. King’s birthday, not to mention Mexican Independence Day, or Kwanzaa, and making things is not the first image to come to mind. The shared sensibilities of family and community stir remembrances of laughter, talk, food, music, and games well before anyone thinks about making a cardboard dreidel or coloring in a Kinara candleabra. Even in schools that acknowledge Halloween, the best aspect is that the real thing is but hours away.

What would it take to “un-Scrooge” school board decision-making apparatuses, and what steps might liberate teachers with holiday policies that are meaningful, culturally relevant, culturally expansive, and, dare I say, fun and worthwhile? The following four principles are offered as a road map to a holiday approach that, if embraced, can (notwithstanding several predictable and necessary bumps, land mines, and hiccups) become potential causes for celebration.

Holiday Principle #1:

Elevate the Involvement of Parents.

Parent involvement is the topic of in-services, classes, study groups, and collaborative research projects. The evidence of positive correlation between parent involvement and student efficacy is indisputable, and most important the literature has demonstrated that empowered and meaningful involvement is the key to a genuine partnership in the classroom (Epstein 1988, 2001; Henderson and Mapp 2002). The problem is that standards-based and commercially produced curricula have limited where parental perspective can actually matter and influence the school day.

Bringing parents into the planning circle is the dimension emphasized in Julie Bisson’s *Celebrate: An Anti-Bias Guide to Enjoying Holidays in Early Childhood* (1997), a trailblazing book that pioneers the inclusion of family perspective at every turn. According to Bisson,

Parents and [g]uardians have critical information that you will need in order to select and implement inclusive, *culturally*

relevant holiday activities. In addition they deserve communication about how their children may be affected by changes in holiday curriculum, and you will need their support in order to successfully put your changes into place. (p. 17)

Bisson, however, wrote for a child-care-center and preschool audience. There is no similar anti-bias holiday curricular publication for the public school teacher at the early grades. The parent survey, a common enrollment document at the kindergarten level in public schools, recedes in frequency as children move up the grade scale. Imagine then the following statement on a survey sent home at the beginning of kindergarten:

We are interested in incorporating holiday observance as part of a culturally relevant, community-building curriculum.

Would you be willing to help select and plan holiday events in your child's classroom?

Teachers must bring to such communication an understanding of many families' historical marginalization in public schools. Often a core group of parents represents the dominant-culture perspective and background. Although European-American families will respond to and take the initiative about such a unique and intriguing invitation, including the diversity of classroom families is essential for the program to have meaning and educational purpose. An inclusive effort can set into motion a new level of adult education and communication that profoundly benefits children.

Out of meetings and task-force planning work could arise an information bank of family tradition and knowledge to help families get to know each other better. This collection of information could also, by deciding which holidays receive special attention, help make more specific plans and integrate each holiday observance into an ongoing range of curriculum beginning well before an actual calendar date, and extending well beyond.

Holiday Principle #2: Integrate Holidays into the Curriculum in a Real and Meaningful Way.

It is often said, "Don't let Thanksgiving be the first time you discuss Native Americans," or "Study black history throughout the year, not just on January 15." Holidays contribute to a trivialized, "tourist" curriculum if not contextualized in consistent, in-depth curriculum development and planning. A holiday has authentic meaning only when presented as a special day in the everyday lives of ordinary people who share both cultural differences and commonalities.

Parent-teacher planning collaboration permits the time and resource gathering that can bring such integration to fruition. A forward-looking planning team anticipating Cinco de Mayo during preliminary meeting time in September might consider this organizing question: What curricular steps include the Mexican-American perspective in a naturally unfolding classroom experience? The answer might involve examining children's literature used in the class; music and songs; stories told; food and snacks shared; and honoring what children and families have to say and offer so that this cultural identity becomes an everyday presence, along with many other group identities represented (and not represented) in class demographics. Of profound importance is the resonation of at least a few languages other than English, well beyond counting to ten and the translation of *My Little Teacup*—even in monolingual classrooms.

Holiday Principle #3: Go Beyond Make'n'Take.

One reason American teachers are wedded to look-alike craft projects and single-line captions (for example, a stove pipe hat with an Abe Lincoln fact penciled in) is the individualism their pedagogical structure emphasizes. Learning becomes the production of work outputs varying only slightly but producing one-per-customer bulletin board displays. That orientation precludes, for example, even considering a mural produced by a group of five, or a skit written by a team, or a ritual of giving thanks that includes (heaven forbid!) a whole-class expression of gratitude for an agreed-upon, community-articulated belief. One of the first questions a teacher can ask is: What are the music stories connected with this holiday? For example, in *Raising Nuestros Niños: Bringing Up Latino Children in a Bicultural World* (1999), a treatise on the culturally grounded development of Latino children, Gloria Rodriguez shares several Mother's Day (*El Día de la Madre*) songs; learned and sung together at rug or classroom meeting time, they can provide a foundation for community building, diversity as a normalized experience, and immunity from the strictest school board admonitions (or even resistant jingoistic ideology).

Even the 100th Day of School has re-introduced games to the celebration matrix. Lacking in the scenario, though, is cultural or family tradition, such as dreidel or Zuni and Hopi stone games: that is, competitions that can be adjusted so that the objective of winning "A Hundred" remains. Such inclusion sets context and continuity for infusing the activities during a holiday such as Hanukkah, or when considering the Native American perspective. At the same time, thoughtful teachers can research cooperative games with links to holiday lore and tradition.

Cooking (and eating) calls for engagement, vocabulary, receptive and expressive language, mathematical application, scientific observation, multiple-intelligence exercise, and literal immersion in environmental

print. Almost every holiday has a food dimension. To acknowledge Thanksgiving without kitchens is akin to teaching handwriting without pencils.

Holiday Principle #4: When It Comes to Holidays, Let Anti-Bias Curriculum Be the Rule.

Rare is the holiday that lacks an anti-bias dimension. The job of the classroom teacher begins with the decision about which holidays to acknowledge, and which to ignore. For example, an anti-bias opportunity is lost when International Women's Day (March 8) passes without mention. Thanksgiving is unavoidably a mythological master-narrative event. The primary-classroom teacher is confronted with opportunity and choice if any of that history is to be challenged, broadened, and demythologized. At the most rudimentary level, Wampanoag, the tribal name of the Native American participants in the November 1621 event that came to be a legend of friendship and community, can be shared and spoken. I was forty-three years old before I learned their name.

Many holidays are inexorably connected to themes of justice and freedom, palpable when considering Passover, Kwanzaa, Independence Day, Veteran's Day, Cinco de Mayo—and the list goes on. Repeatedly, anti-bias literature has demonstrated young children's sensitivity to stereotyping and other falsehoods based on group membership. The *Anti-bias Curriculum* video produced by Pacific Oaks College (1988) portrays a striking scene: four-year-olds critically deconstructing a collection of holiday cards filled with stereotypes.

This kind of critical thinking tends to complicate and confuse bits of knowledge once thought intractable. Such a process can be disquieting to the adult world. Our dominant culture has a tendency toward ego- and sociocentricity about such matters. In child-development nomenclature, disrupting this paradigm is referred to as *disequibration*: the process of previously held knowledge becoming stretched, bent, elaborated, complicated, differentiated, and more veritable. Throwing children's knowledge about holidays into disequilibrium and then reconstituting it at more-sophisticated levels advances their cognitive, social, and emotional skills. A teacher who helps facilitate that transformation would be justified in saying that holidays have provided incomparable learning opportunities.

References

- Bisson, J. 1997. *Celebrate: An Anti-Bias Guide to Enjoying Holidays in Early Childhood Programs*. St. Paul, Minn.: Redleaf Press.
- Derman-Sparks, L. 1988. *Anti-Bias Curriculum*. Video Kit-Teacher Reference. Pasadena, Calif.: Pacific Oaks Bookstore (POB. 30 min.).

- Epstein, J. 1988. "How Do We Improve Programs for Parent Involvement?" *Educational Horizons* 66(2): 58--59.
- . 2001. *School, Family, and Community Partnerships: Preparing Educators and Improving Schools*. Boulder, Colo.: Westview Press.
- Henderson, A., and K. Mapp. 2002. *A New Wave of Evidence: The Impact of School, Family, and Community Connections on Student Achievement*. Austin, Tex.: Southwest Educational Development Laboratory.
- Phillips, C. 1989. *Nurturing Diversity for Today's Children*. Young Children. NAEYC.
- Rodriguez, G. 1999. *Raising Nuestros Niños: Bringing Up Latino Children in a Bicultural World*. New York: Fireside.

Cory Gann is an associate professor in the College of Education and Professional Studies at Central Washington University, Des Moines, Washington.

Characteristics of an Effective Student Testing System

by Richard P. Phelps

The U.S. public has consistently favored standardized testing in the schools, preferably with consequences (or “stakes”) riding on the results, ever since the first polls taken on the topic several decades ago. Depending on how the question is framed, those in favor of high-stakes standardized testing outnumber those opposed at ratios as high as twelve to one. Parents are stronger supporters of high-stakes testing than are nonparents, and that support does not budge when they consider the possibility of their own progeny failing.

Results from different polls approaching the topic in different ways suggest that nearly all Americans would like to see high-stakes tests administered at least once at every grade level. In twelve years of elementary and secondary school, however, the typical U.S. school district offers just one or two standardized tests with high stakes for students.

With only a few exceptions, U.S. educational testing programs fall short of what the public wants, and short of what most industrialized countries have.¹

Comprehensive examination systems are multi-level and multi-targeted

A comprehensive testing system captures the complete benefits of standardized testing—and for all students, not just some. Those benefits include:

- *Information* that can be used for diagnosis (of individual students or teachers, of schools, of school programs)
- *Efficiencies* from alignment, when the tests are matched to curricular standards and teachers teach to those standards (and yes, teach “to the test,” as they are supposed to do with standards-based tests)
- *Motivation* to study and to attain goals

The best testing regimes, such as one finds in many European and Asian countries, capture those benefits through multi-level and multi-target systems.

Multi-level means administering high-stakes tests at more than just one educational level (i.e., primary, intermediate, lower and upper secondary). European and Asian students typically face high-stakes tests at the beginning or the end (or both) of one educational level, and often for more than one educational level (e.g., the end of primary school, the beginning and end of lower- and upper-secondary school, the beginning of postsecondary education, etc.).

Multi-target means that every student, regardless of achievement or choice of curriculum, faces a high-stakes test that, ideally, offers a challenging but attainable goal. In some systems, tests are set at multiple levels of difficulty with multiple levels of certification (e.g., a “regular” diploma and an “honors” diploma). In other systems, different tests cover different subject matter.

In the United States, high-stakes student tests are uncommon at any but the upper-secondary level (i.e., high school). Moreover, with few exceptions, they are single-target tests—meaning that every student, no matter what level of achievement or ability, course selection, or curricular preference, must meet only one common standard of performance.²

Ironically, largely socialist Europe, with its relatively smaller socioeconomic (and academic achievement) disparities, acknowledges children’s differences by offering a range of academic options and multiple achievement targets. The more libertarian United States, with its relatively large socioeconomic (and academic achievement) disparities, nonetheless typically provides all children the same curriculum—often called the “college track”—and sets a single academic-achievement target.

A single academic-achievement target must by necessity be low: otherwise, politically unacceptable numbers of students will fail to reach the targets. School systems that set low targets typically focus on bringing the lowest-achieving students up to that target. Unfortunately, they may also neglect average- and higher-achieving students (or, in the most perverse cases, deliberately hold them back). Schools judged as a whole on student performance can increase their average scores, for instance, by retaining high-achieving students with their age-level peers rather than letting them advance a grade or by making them take courses in basic subject matter they have already mastered.

The differential effect may help explain why some minority-rights advocates support minimum-competency testing, while parents of “gifted and talented” children often oppose it. From the perspective of

The “High Stakes Cause Test-Score Inflation” Myth

John J. Cannell’s late-1980s “Lake Wobegon” reports suggested widespread, deliberate educator manipulation of norm-referenced standardized test administrations and scores, resulting in artificial test-score gains—such that every U.S. state had “above average” test scores. The Cannell studies have consistently been referenced in education research since, usually as evidence that high stakes, not cheating or lax test security, cause test-score inflation—this despite the fact that only one of the dozens of Cannell’s score-inflated tests had stakes attached.

In fact, a careful reading of Cannell’s reports shows that low, not high, stakes are associated with test-score inflation. Low-stakes tests make cheating possible because those tests are often administered with lax or no security.

Conversely, high-stakes tests are more likely to produce reliable test results because those tests are typically administered with tighter security. Given current law and practice, the typical high-stakes test is virtually certain to be accompanied by item rotation, sealed packets, monitoring by external proctors, and the other test-security measures itemized as necessary by Cannell in his late-1980s appeal to clean up the rampant corruption in educational testing and reporting.

Other test-security enhancements that also tend to accompany high-stakes tests include a high public profile, media attention, and voluntary insider (be it student, parent, or educator) surveillance and reporting of cheating. Do a Web search of stories about test cheating, and one finds that, in many cases, cheating teachers were turned in by colleagues, students, or parents. Public attention does not induce otherwise honest educators to cheat; it enables otherwise successful cheaters to be caught.

—R.P.P.

See Phelps, R. P. 2005. “The Source of Lake Wobegon.” *Nonpartisan Education Review* 1(2). <<http://www.npe.ednews.org/Review/Articles/v1n2.htm>>.

the former, the tests pull achievement levels up. From the perspective of the latter, the tests pull achievement levels down.

The single-target problem has two solutions, one passive and one active. The passive solution, currently used in many U.S. states, essentially involves letting individual students take the minimum-competency test early in their school careers; once they pass it, they are allowed to move on. If the test is high stakes only for individual students, no one has an incentive to hold higher-achieving students back, that is, to prevent them from taking accelerated course work from then on.

The active solution to the single-target problem, and the solution that promises greater overall benefits, is to offer multiple targets. New York has stood out historically as the one U.S. state that employs a multiple-target examination system, with its Regents "Competency" exams and Regents "Honors" exams. The former was required for high school graduation with a "regular" diploma, while the latter was required for graduation with an "honors" diploma.

European and Asian examination systems exist in a variety that reflects the educational programs offered. Students are differentiated by curricular emphasis and ability level, and so are their high-stakes examinations. The differentiation, which starts at the lower-secondary (i.e., middle school) level in many countries, exists in virtually all of them by the upper-secondary level. Students attend schools with vastly different occupational orientations: advanced academic schools to prepare for university; general schools, for the working world or for advanced technical training; and vocational-technical schools, for direct entry into a skilled trade. Typically, all three types of school require an exit examination for a diploma. Some of those exams are very tough.

Supporters of the one-size-fits-all U.S. system often label the European system as "elitist" and our system as a more "democratic," "second chance" system. That contrast may have been valid forty years ago, but no longer. It is easier to enter upper academic levels in the current European systems, and most countries now offer bridge programs for, say, a dissatisfied vocational-track graduate to enter university or advanced technical programs. Typically, bridge programs are free of charge.

If the U.S. system is neither less elitist nor more conducive to "second chances," how is it superior? It is not, really. In the typical European system, multiple programs and multiple tracks offer multiple opportunities for students to attain high achievement levels. A Swiss, German, Danish, or Austrian student who enters a vocational-technical track at the lower-secondary level and finishes by passing the industry-guild certification examination as a machinist enters an elite of the world's most skilled (and best-paid) craftspersons. By contrast, a vocational-technical student in the United States may be stigmatized by a curriculum with a

reputation as a “dumping ground” and receive only low-quality training, with out-of-date equipment, for low-level jobs.

Fair high-stakes tests are aligned to standards

Most high-stakes student examinations are aligned to common standards; it is simply not fair to attach stakes to a test containing content to which students have not been exposed. What is more, no standards-based test, regardless of the care and effort put into writing it, can salvage poor curricular standards.

Profound disputes over curriculum and instruction are the major reasons high-stakes state tests can vary so widely in character. Take the neighboring states of Maryland and Virginia, for example. Several years ago, Maryland’s School Performance Assessment Program (MSPAP) incorporated test administrations at three different grade levels, and performance carried consequences for schools. The entirely “performance-based” test had no multiple-choice test items and even included group work and “hands-on” demonstrations. It emphasized “process” over content. By contrast, Virginia’s traditionally administered, content-oriented Standards of Learning examinations (SOLs) contained a large proportion of multiple-choice items, completed in their entirety by individual students.

Different theories about what should be taught and how it should be taught underlay the development of the Maryland and Virginia examination programs. To be sure, different theories of assessment were also involved, but they were inextricably tied to curricular and instructional preferences. Only the most extreme testing opponents decried both the Maryland and Virginia tests. Many “progressive” educators liked Maryland’s, whereas many “traditionalists” preferred Virginia’s.

Examination systems require careful implementation

Even if one assumes that, say, the French examination system is worth emulating, a U.S. state with no testing program cannot replicate it overnight. The French system is supported by a relatively uniform curriculum-development system, which is managed by university subject-area experts. This developed curriculum buttresses several (or many, in vocational areas) curricular tracks that students can follow. Students are provided multiple opportunities to pass the examination of their choice, and they receive substantial help, such as further classes and tutoring, to pass those examinations. But they must pass before they can go on to university, polytechnic, or a specialized trade. Although given every reasonable aid to succeed, in the end they must know the basic subject matter of their chosen path, or they will not be allowed to proceed at taxpayers’ expense.

Two general issues are involved in building an examination system: sequencing and structuring.

Sequencing is the more straightforward of the two. Implementing a standards-based test requires time and care. The standards must first be in place, and taught to, before students can be tested fairly. New tests then should be field tested to address problems and set baselines for performance. The process takes at least a year, and more commonly two or three.

Most U.S. states building new examination systems have started with tests designed to measure minimum levels of knowledge and skill deemed adequate for earning a high school diploma. Passing a minimum-competency examination can frequently imply nothing more than a sixth-, seventh-, or eighth-grade level of achievement.

A minimum-competency exam brings a state only to the edge of the French examination system, however. There, minimum-competency examinations are given at the end of lower-secondary school; passage is required before a student can move on to upper-secondary or specialized vocational schools. The students who advance through this next level of education in France choose a curricular track and then face tough exit examinations of a type and level of difficulty that scarcely exist in the United States.³

Examination difficulty, in fact, introduces another aspect of sequencing. Some U.S. states have constructed high-quality examinations aligned to their standards only to discover, during field tests, that few students could pass them. Any state in which the majority of students fails a test required for graduation—only a year after the untested students of the previous class all graduated—will face a public uproar.

Aside from merely easing the difficulty of the examination, the problem has at least two feasible solutions. One can start easy and gradually ratchet up the difficulty level of the examination, or one can provide students extra assistance to pass the examination. The latter strategy was adopted with great success in Massachusetts. The state's elected officials, however, absorbed considerable invective from testing opponents (including one state teachers union) while they stood firm on the standards. The Massachusetts strategy is not for the faint of heart.

Structuring is more varied and complicated than sequencing. In countries with well-developed testing systems, two general types are distinguishable by relative degree of curricular specialization, or "splintering."

- The French example above describes considerable curricular splintering or tracking, common to the European "continental" system. Starting at the beginning of lower-secondary school, or perhaps even earlier, students are tracked into different types of

schools according to ability level and personal and parental choice of curricular focus.

- The traditional “two-tiered” British system represents another examination system: general curricula well into high school, but at two levels of difficulty—the “O,” or ordinary level, and “A,” or advanced level. The former two-level Regents examination system in New York State also represented this model.

Whatever the testing-system model employed, it should make sense as an integrated whole. To be fair to all students, a testing system should offer opportunities and incentives to all students, and students are not all the same.

Consequences of eliminating standardized testing

Standardized testing has often been measured against utopian perfection rather than what is likely to take place in its absence. It is true that neither standardized tests nor the manner in which they are administered will ever be perfect, but the consequences of abandoning standardized testing are far from perfect, too.

One likely consequence of eliminating standardized testing is a system of social promotion with many levels of nominally the same subject matter, ranging from classes for the self-motivated kids to those for the kids who quit trying years before, kids the system has ignored ever since. Too often, the result is a system that graduates functional illiterates. In schools where students are routinely passed whether or not they earn it, teachers brave enough to assign failing grades may well have their marks erased and changed by school administrators, thus allowing the failing students to graduate and avoiding controversy. In schools where some students pass courses and graduate despite doing little work, other students, and their parents, will assume that they, too, can pressure school administrators for easy credentials. Behind-the-scenes prerogatives become the implicit academic standards.

Another likely consequence of eliminating high-stakes standardized testing is the large-scale institution of remedial programs in colleges to compensate for any deficiencies of instruction in elementary and secondary schools.

A third likely consequence of eliminating high-stakes standardized testing is a blackout of reliable information on student performance anywhere outside a student’s own school district. Eliminating *high-stakes* standardized testing would increase schools’ reliance on teacher grading and testing, which are far more likely to prove idiosyncratic and non-generalizable than any standardized test. Individual teachers can narrow the curriculum to what they personally prefer. Grades are susceptible to

inflation as students learn teachers' idiosyncrasies and how to manipulate their opinions. According to research on the topic, many teachers, when assigning marks, tend to consider noncognitive outcomes, including student class participation, perceived effort, progress over the period of the course, and comportment. Actual subject-matter mastery is just one among many factors. Moreover, given most teachers' relatively brief training in testing and measurement, it is not clear that their testing and grading practices would be superior even if they focused only on subject-matter mastery.

If the curriculum is not tested, it is difficult to know if any of it works. Without standardized tests, reliably gauging student progress becomes problematic for anyone outside the classroom. One must accept whatever each teacher says, and without standardized tests, points of comparison for different classrooms become progressively rarer.

Without either common standards or high-stakes standardized tests, there may be no effective way to monitor systemwide performance *at all*. Some U.S. teachers may be doing a wonderful job in their totally customized classes, but some may be doing an awful job. How is one to know or tell which? One must hope that teachers will face down their own natural inclinations as well as those of students, parents, and schools to avoid accountability and hold themselves and their students to high standards of performance regardless. One must also hope that teachers will know how.

This document excerpts from Kill the Messenger: The War against Standardized Testing, by Richard P. Phelps, published by Transaction Publishers. Copyright © 2003 by Transaction Publishers, Rutgers University, 34 Berrue Circle, Piscataway, NJ 08854. ISBN 0-7658-0178-7 / cloth / 331 pp.

Richard P. Phelps, the author of Kill the Messenger, upon which this article draws in part, is also the editor of Defending Standardized Testing (Lawrence Erlbaum 2005) and Correcting Fallacies about Educational and Psychological Testing (APA Books 2007), and a member of the editorial board of the Nonpartisan Education Review (www.nonpartisaneducation.org).

Tests on Trial		
<i>Judging standardized tests against a benchmark of utopian perfection that does not and cannot exist means standardized tests always look bad. How would the criticisms look compared to the actual, available alternatives?</i>		
The Case Against Standards and Tests	The Testing Rebuttal	What's More . . .
Teaching to the Test		
<i>Teachers will teach only material that will appear on a standardized test.</i>	If high-stakes tests are kept behind lock and key until the day of test administration, teachers will not know what material will be on the test, except in the most general terms.	In the absence of common standards and tests, the curriculum becomes arbitrary and of uncertain origin. Why is that better than teaching to a required curriculum.
Narrowing the Curriculum		
<i>A common curriculum prescribed by standards has less content than a teacher-made curriculum.</i>	A school year's fixed number of hours and days renders it unlikely that a common curriculum will have less content than a teacher-arbitrary curriculum. I.e., a teacher who drops one topic when standards are introduced has necessarily added another.	What teachers and schools do in the classroom without common standards is not necessarily "broader." In fact, it can often be "narrower"—governed in the absence of other criteria by personal preferences.
Cheating by Students		
<i>High-stakes standardized testing increases students' incentives to cheat.</i>	Cheating is far easier to prevent and detect with standardized tests. Different forms used in the same classroom can make copying unrewarding. Computer programs run after the fact can look for telltale patterns.	Cheating in regular classroom work has become epidemic. The overwhelming majority of students admit to cheating in polls. Teachers and schools are ill equipped to monitor or detect most cheating. Meanwhile, the Internet makes cheating far easier than it used to be.
<i>(continued)</i>		

Cheating by Teachers		
<i>Many teachers have been caught cheating on high-stakes standardized tests.</i>	The fact of detection may be evidence of how easily such cheating can be detected.	Social promotion and grade inflation provide evidence that nonstandardized testing and grading are far from infallible. And consider that in surveys, the majority of teachers claims overwhelming pressure to award high grades to undeserving students.
Preferred Instructional Methods		
<i>Classrooms governed by standards are barren, dreary places where only factoids are learned.</i>	A curriculum will always rely on some sort of standard or criteria for inclusion. The question is, Do we want formal, open standards, openly arrived at, or should their origins be more obscure?	Many teachers, especially inexperienced or quick hires in underperforming schools, will rely on the teacher's versions of basal textbooks for course content or, worse, make it up. Is the classroom shorn of standardized testing automatically a wonderful place—rich with innovative curriculum, the joy and magic of learning, and so on? What is the evidence?
Opposition to Norm-Referenced Tests		
<i>Norm-referenced standardized tests are unfair. (I.e., it is unfair to simply rank kids, rather than measure them against standards.)</i>	Norm-referenced tests provide information that cannot be obtained any other way. Many educators find them useful as measurement benchmarks and for curricular diagnosis.	The alternative, grade-point averages, are norm-referenced measures, normed at the school level.

Preference for Teacher-Made Classroom Testing		
<i>Standardized tests are imposed from outside by persons and committees unfamiliar with and perhaps insensitive to the local students and community.</i>	Standardized tests are developed by testing and measurement Ph.D.'s. The most capable measurement experts in the world work in North America developing standardized tests.	Teacher ed programs provide few teachers with more than cursory training in measurement, yet the absence of standardized testing would have us rely exclusively on their measurement decisions.

Notes

1. For a roundup of such polling results, see my “Persistently Positive: Forty Years of Public Opinion on Standardized Testing,” chapter 1 in *Defending Standardized Testing*, ed. Richard P. Phelps (Mahwah, N.J.: Lawrence Erlbaum, 2005), 1–22.

2. The federal No Child Left Behind (NCLB) Act set in place what is largely a testing program. NCLB, however, falls far short of a comprehensive multi-level, multi-target high-stakes testing system; it sets only one achievement target (for schools), establishes no stakes for students (little motivation to take the test seriously), and provides curricular alignment that can be less than perfect.

3. An ambitious American student could simulate an equivalent program by taking several Advanced Placement (AP) examinations, except that she will graduate from high school and be accepted by some college no matter how she scores on them. Some states (e.g., Virginia, Michigan) are attempting to build something like this structure by requiring passage of a certain number of end-of-course examinations in high school.

Perils of Standardized Achievement Testing

by Thomas M. Haladyna

Abstract

This article argues that the validity of standardized achievement test-score interpretation and use is problematic; consequently, confidence and trust in such test scores may often be unwarranted. The problem is particularly severe in high-stakes situations. This essay provides a context for understanding standardized achievement testing, then presents and discusses threats to validity, many of which are currently unaddressed. The public and several constituencies support standardized achievement testing. Many educators, however, especially educators in testing, have argued consistently that test-score interpretations and uses are inadequately validated. Standardized achievement test scores provide one valid source of information about student learning if they corroborate other information about student learning. Unfortunately, so many factors undermine the validity of test scores that we should be very careful in the way we interpret and use them.

Richard Phelps (2006) correctly points out many facts about standardized achievement testing: the public wants it, other nations may do it better, and many critics offer no viable alternatives to it. As he concedes, though, standardized achievement tests will never be perfect. Given this state of affairs, what should we do about standardized achievement tests in America? Do we reject the messenger of student achievement? Should we have higher standards for tests when the outcomes of test scores are high stakes? Should we exercise some caution when we use such test scores?

The thesis of this article is that using standardized achievement test scores for high-stakes purposes is perilous because many threats to test validity have gone unaddressed. Part I of the article provides a context

for discussing these perils. Part II identifies and discusses the ways in which such perils threaten validity. Part III then attempts to answer the questions stated in the first paragraph.

PART I: THE CONTEXT

The aims of Part I are explaining basic concepts; describing what students learn; presenting a model for student learning; discussing the role of testing in this model; and finally, considering the role of validity and validation in standardized achievement testing.

Basic Terms

The jargon that permeates education constitutes one difficulty in discussing standardized achievement testing and student learning. Several terms are defined here.

Student *achievement*, distinguished from *intelligence*, is cognitive behavior changed by learning experiences. Intelligence and the cognitive abilities that make up intelligence are less subject to such change. Generally, achievement and intelligence are highly correlated. In most states a set of content standards, which reflect what students should know and can do, defines student achievement.



Test. A test is a measuring instrument. An achievement test, if validated, measures student achievement. Without validation, it is hard to make and justify a claim that an achievement test measures student achievement.

High-stakes test. Some uses of test scores have significant consequences for students, teachers, schools, and school districts. The term “high stakes” designates a test-score use with such consequences, including graduation or promotion; school accountability, such as the federal No Child Left Behind (NCLB) legislation requires; merit pay or continued employment based on test scores in schools or school districts; and intervention in schools or school districts due to chronically low achievement test scores.

A teacher undertakes *assessment* after collecting information about student learning. Assessment, a judgment about student learning, helps plan future instruction. A standardized achievement test score is one piece of information useful for the assessment. We often mistakenly equate the terms “assessment” and “test,” but in this context the test is clearly what we use to help us make an assessment.

Accountability. In the past, accountability meant providing information to policymakers to aid their decisions about instructional programs and resources for students. A newer interpretation of accountability—holding people responsible for student learning—is simplistic in its logic: teachers are seldom fully equipped to deal with student learning, and they often have little control over resources needed to help students learn. In test-based accountability, the test score becomes the only basis for assessing a student or a group of students in a classroom, school, school district, or state. One test, however, should never be the sole basis of assessment (AERA 2000); other information should corroborate the test score and better inform us about student learning.

Validity. Validity refers to the adequacy of any test-score interpretation. Let’s say Bob, the fastest runner in our high school as a freshman, refused to run hard during a time trial. He finished a 100-yard dash in 15.2 seconds. In other words, his standardized test score was 15.2. Would the coach’s assessment of his running speed based on that result be valid?

What Students Learn

A graduate of the University of California at Berkeley has stated that what got him through “Cal” in the 1950s were the three Rs: read, remember, and regurgitate. What he regurgitated was knowledge at the lowest cognitive level: recall. Most of us have experienced that kind of learning. Recall is still part of learning, but understanding and using knowledge are now also widely recognized as dimensions of student learning. Both

can be conceived of as domains that contain many tasks. A test is a sample of the tasks from a domain.

The first domain of student learning consists of *knowledge* and *skills*. Knowledge exists as facts, concepts, principles, or procedures. Knowledge can be recalled, understood, or applied. Spelling and punctuation are examples of writing skills. Most student learning involves knowledge and skills in the domains of reading, writing, mathematics, science, and social studies. Each subject has a large domain of multiple-choice test items representing knowledge and skills at all levels of proficiency. The multiple-choice format, which has proved effective and efficient in measuring knowledge and skills (Downing and Haladyna 2006; Haladyna 2004), should continue in use as a valid measurement of the knowledge and skills domain.

The second and newer domain of student learning draws from cognitive psychology and the persistent belief that learning entails more than simply learning knowledge and skills. *Cognitive ability* is another name for what this domain represents: a mental capacity for achieving an end through complex use of knowledge and skills. Different writers apply different names to this capacity: *developing ability* (Messick 1984); *fluid ability* (Lohman 1993); and *learned ability* (Sternberg 1998). Each cognitive ability is easily recognizable: reading, writing, speaking, listening, mathematical and scientific problem solving, and critical thinking. The acquired knowledge and skills of the first domain are put to use in this second domain. Tests of cognitive ability require that students apply and not simply regurgitate knowledge. Skills are used in unique and complex ways. Standardized achievement tests were not designed to measure cognitive abilities. The advent of state content standards motivated by the federal No Child Left Behind legislation makes clear that future tests will have to address how we use knowledge and skills. Students will be learning knowledge and skills they can apply in their own lives. The performance-test format is best suited for measuring cognitive abilities such as writing.

A Model for Student Learning

Carroll's classic generic model for student learning (1963) holds that educators present their students with clear learning outcomes; aligned instruction; aligned measurement of student learning that provides a valid basis for assessment; and re-instruction where needed to achieve student-learning goals. This model for student learning has not been altered through the years, but the standards for aligned instruction and testing have greatly improved. Not only must today's standardized achievement test be aligned to our content standards, but the alignment of instruction both to content standards and to assessment tests must

also be demonstrated. All students must be provided opportunities for learning and re-learning until they meet desirable performance levels. Both NCLB and AERA (2000) promote this model for student learning and provide guidelines.

The Role of Testing in This Model

Two main uses of test scores are helping teachers improve future instruction by assessing student learning, and providing responsible parties (i.e., both district and state school boards; state and federal legislators; and the public) with information about student learning. Many of those constituencies need such information to formulate policy and allocate resources to schools. In some circumstances, accountability use includes graduation or promotion testing.

Any and all uses of test information must be validated (AERA, APA, and NCME 1999). Without validation by a test-score interpretation, the information culled from a standardized achievement test is dubious.

Validity and Validation

Messick (1984, 1989, 1994, 1995a, 1995b) and Kane (in press) identify validity as the most important goal in testing. Validation is the investigative process that appraises validity for test-score interpretation. *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999) is clear about how to validate test-score uses. The process of validation has many steps: first, defining the content being tested; then proposing the interpretation of a test score (the test developer must argue that the test is created to measure this trait validly); and later, gathering evidence to support the claim of validity (Haladyna and Olsen 2006). Validation is a long-term study of a test; the goal is to improve validity.

PART II: PERILS OF STANDARDIZED ACHIEVEMENT TESTING

Haladyna and Downing (2004) show that the many factors threatening validity fall into two main categories: 1) content irrelevance—factors that incorrectly and systematically increase or decrease test scores for some students, and 2) content underrepresentation—flaws in the design of the test that fail to evaluate its full range of content and cognitive behaviors.

Those factors weaken, undermine, or destroy validity. One obvious factor is cheating, which inflates a test score inaccurately. We need to investigate standardized achievement test scores before endorsing and accepting them as unquestionable truth. By studying each factor, we can

reduce or eliminate a threat before using test scores as desired. Eliminating or reducing threats increases score validity. Presenting this information to the public should provide proof that test results can be trusted.

The following discussion focuses on high-stakes uses of standardized achievement test scores. (With low-stakes uses, the need for validity is important, but not to the same extent.) The relevant factors include

- students
- instruction
- test preparation
- cheating
- test development
- test administration
- test scoring
- standard setting

Some of these factors are more serious than others, but all undermine validity to some extent and all have been documented in American testing.

Students

Students themselves are one major source of contamination in testing. Students who cannot read the test material adequately tend to stop taking the test or to mark answers aimlessly. Our inference may be that they have not learned, when a more fundamental problem exists: they cannot read. Nonresponse and omitted responses are more prevalent with English language learners (Haladyna, Osborn Popp, and Weiss 2005). Other factors seldom assessed when students take standardized achievement tests include motivation and fatigue; varying incentives among schools and school districts for performance on the test (Haladyna, Nolen, and Haas 1991); and the motivation level of different students—some students are highly motivated, whereas others seem not to care. Random marks or large blocks of unanswered items are scored as “wrong” when they were in fact omitted from the test by the student. The resulting test scores are inaccurate.

Instruction

NCLB and the new version of accountability have ensured that teachers align instruction to the state’s curriculum. The transition from a relaxed, enlightened selection of subject matter to a more rigid system requiring grade-appropriate goals in each subject matter can fulfill the model of student learning presented earlier in this article. Other nations’

unified curricula may make their alignment and testing more uniform and effective than does the United States (Phelps 2006), but most U.S. states have abandoned “states’ rights” to follow curricular guidance provided by national organizations.

Opportunity to Learn (OTL). An outgrowth of the accountability movement has been nearly universal support for OTL. AERA and the National Council of Teachers of English, among many organizations, have espoused OTL standards that address various concerns: content standards-based instruction; the diverse ways students learn; highly qualified teachers; best classroom practices; and assessing schools and classroom learning environments. Unfortunately, few state or school district standardized-testing programs assess the conditions of classroom learning that directly address OTL. Despite accountability’s premise of adequate instruction for all students, insufficient information about OTL makes it nearly impossible to evaluate student learning for high-stakes purposes such as graduation. In other words, students must first be taught before we use a test to measure what they should have learned. Further, students who are not initially successful should be given repeated opportunities to learn so that they really are not left behind.

Lack of Test Alignment. The introduction of the Stanford Achievement Test in 1923 initiated widespread evaluation of student learning by such tests, whose alignment with an idealized curriculum was regarded as an advantage. However, the trend toward state content standards has caused many states to abandon the publishers’ tests in favor of tests aligned to the new standards. The publishers’ tests have survived because they provide norm-referenced comparisons for millions of students. Nonetheless, norm-referenced test-score data can be vexing. Haladyna (2004) pointed out that the National Assessment of Educational Progress (NAEP) ranked one state’s students fourth from the bottom nationally, whereas they performed well above average in a nationally normed standardized achievement test. One suspicion is that the latter test was the object of considerable coaching; thus the national norms were compromised and state policymakers were misinformed about their students. This tendency, originally called the “Lake Wobegon effect” in reference to humorist Garrison Keillor’s mythical Minnesota town where all the children are above average, is still prevalent today.

Test Preparation

In my files is the Stan Fordnine test, a cloned version of the Stanford 9, used for test preparation in one high-scoring school district. When teachers can study a test and identify the content and specific objectives that each test item measures, there is a strong temptation to teach content that will directly affect test performance. This practice is known

pejoratively as “teaching to the test.” Advocates of standardized achievement testing and test-based accountability often defend teaching to the test, claiming it is better to learn something valued than not. Teaching to the test, however, is a type of consumer fraud. Any test is only a sample from a large domain of knowledge and skills; mastering a small part of the domain that happens to be tested creates a biased test score. Students, parents, and the public think that more learning has occurred than really did.

Test preparation must be ethical (Haladyna, Nolen, and Haas 1991). Many well-documented test-preparation practices essentially “trick” test scores for the obvious advantage of making it appear that more learning has happened than was the case. The most ethical test preparation is good teaching: using the content standards, aligning instruction to these standards, assessing learning, re-teaching, and re-assessing. The teacher follows all content standards for the grade level and aligns instruction and assessment with that content. Teaching to the test is one possibility when test scores increase in peculiar patterns from one year to another. The tactic may work in test-based accountability, but only in the sense that the public is fooled.

Cheating

Cheating on tests is a pervasive problem in American education as well as throughout the world. Test scores can be badly corrupted by cheating. A Google search on the Web will yield eighty-two pages of hits on test cheating. The National Center for Fair and Open Testing (www.fairtest.org) regularly reports on cheating, and Caveon, a test-security company, provides weekly updates on test-cheating scandals worldwide (www.caveon.com). That company’s recent survey of thirty-four states (Sorenson 2006) reveals that test security is a great concern: efforts to increase security are on the rise, detection methods are increasing, and more security measures are being planned. Lost or stolen booklets seem to cause the greatest concern.

Recent reports from New Jersey and Texas (www.philly.com; www.npr.org) document the most pernicious problem: educators cheating to satisfy accountability requirements. Without security audits and studies in states and school districts, cheating may go undetected unless a local newspaper investigates and exposes the problem.

Test Development

Test development is a science with a considerable technology. The *Handbook of Test Development* (Downing and Haladyna 2006) provides many instances of standards that apply to test development. Each standard represents an important source of *validity evidence*. Without this

evidence, the validity of any test-score interpretation is in question. In each step of test development, peril exists.

One of the most important categories of validity evidence is content. Is the test content matched to the intended content? Is the test content trivial or learned by rote—or does the test content ask students to apply knowledge in new situations they may encounter in life? Tests can be highly biased samples of what students need to learn. We need assurance and evidence that the content of the test is exactly that prescribed by the state.

Reliability is another important category of validity evidence. Only high reliability can maintain confidence in making high-stakes, pass-fail decisions. High reliability is difficult to attain when writing tests (Haladyna and Olsen 2006). Continued use of test scores for life-altering decisions not only runs the risk of prompting legal challenges; it also damages the students who unfairly fail. One remedy is to ensure high reliability and a small margin of error.

The quality of test items is another important category of validity evidence. Item development is a costly and lengthy process, an estimated \$300 to \$1,200 for a single test item, depending on the effort spent developing an item bank. Evidence that quality of test items is a high priority should be assembled and posted on state Web sites with technical documentation establishing item development and validation. Such evidence is in short supply, as a search of such sites will confirm.

In high-stakes testing, the cut score is a point on the test-score scale at which students are classified one way or another. One important cut-score determination is pass-fail. Another is using a test score to classify each student in one of four categories: highly proficient, proficient, approaching proficient, and well-below proficiency. The test-score scale should mean the same from year to year, because cut scores are standardized for pass-fail decisions and those other classifications. The goal is accurately measuring progress in student learning over several years. The science of scaling for comparability is well established for multiple-choice tests, but not for performance tests (Kolen and Brennan 2004). Scaling for performance tests can be particularly troublesome when tracing growth vertically across grade levels (Haladyna and Olsen 2006).

This discussion is not intended to suggest that test development is faulty or contributes to lower validity, but annual documentation must assure the public that the validity evidence of such high-stakes tests supports reliance on test scores. Technical reports and other indicators of validity covering this evidence should be abundant and comprehensive. Searching Web sites of major test developers, however, reveals that such evidence is scarce and usually nonexistent. A review of such reports (Ferrara and DeMauro, in press) finds them lacking in many

respects. In summary, considerable peril lurks at each step of test development. Without documentation, there is no way to determine what has happened.

Test Administration

Altering the administration of high-stakes standardized achievement tests renders any “standardized” achievement test less standardized. According to anecdotal reports, two common practices are altering the administration time of high-stakes tests for the advantage of the students and reading test items to students. Evidence of such problems in test administration (Haladyna and Downing 2004; McCallin 2006), although growing, is hard to come by because few test sponsors, states, and school districts consistently monitor test administration with care. The most decisive strategy to combat the peril would be hiring professional test administrators. Computer-based tests offer hope of standardizing test administration and reducing this category’s threat to validity.

Test Scoring

A serious source of invalidity is test scoring. The National Center for Fair and Open Testing and other sources, mainly in the media, have found that scoring errors, not to mention monitors sanitizing answer sheets by reviewing them and cleaning up bad erasures, are not unusual. Such practices are unfair when some schools and districts employ them and others don’t.

The scoring of performance tests that require subjective judgments poses many threats to validity, which typically go unacknowledged (Haladyna and Olsen 2006). Results close to the cut score should be re-scored when the stakes are very high, as in graduation or promotion.

The most noteworthy scoring error to date occurred with the Scholastic Assessment Test, the highly respected college-admission test of the Educational Testing Service, one of America’s foremost test companies (National Center for Fair and Open Testing, May 2006): excessive moisture in answer sheets resulted in misscoring. Whether the problem has occurred in the past is unknown. ETS also had to pay \$11.1 million to settle a lawsuit involving 4,100 teachers who received erroneous scores on a teacher-licensing test.

Such incidents are not limited to ETS; the problem of test-scoring errors is widespread. If a test score does not seem representative of a student’s true ability, scoring error is a potential culprit. The major point is that using a single, flawed score for a high-stakes purpose, such as college admission or licensing a teacher, has legal consequences. We need to be smarter in quality control and use of test scores when such errors seem more likely than ever.

Standard Setting

As discussed earlier, high-stakes tests often employ a cut score to designate students as passing or failing. Most commonly, a committee of subject-matter experts reviews items and makes judgments that are aggregated to form the recommended cut score. Another committee may accept the recommendation or change it. Do different methods of development produce different cut scores? Is one cut score more valid than another? Which is the most valid? How does one decide? Those are imponderable questions. Cut scores are set and decisions are made based on arbitrary criteria; the labels used to identify students or groups of students (e.g., proficient, basic) are social conventions, not true categories. The difficulty of tests and the position of the cut scores are sure to vary from state to state. Studies on the validity of cut scores are typically not reported. Consequently, we have very little information about the validity of cut scores used for high-stakes purposes.

Consequences

The benefits and deficits of a high-stakes testing program are its consequences. AERA (2000) argues that negative consequences of test-score uses should be made public. A recent study by Warren, Jenkins, and Kulick (2006) of high-stakes testing's impact on state graduation rates from 1975 to 2002 is not positive. Graduation rates are negatively correlated with high-stakes testing, and the rate of GED testing is increasing in high-stakes states. Although other factors may contribute, those findings support a plausible hypothesis that high-stakes uses of the tests have negative consequences. The public needs to be informed about any negative consequences of standardized achievement testing.

Summary

Threats to the validity of high-stakes standardized achievement test scores are well documented in many sources, both scholarly and popular. This section has sought to point out the often-questionable validity of standardized achievement test score interpretations. Documentation that could assure critics and the public of the scores' accuracy is scarce; in many instances scores are simply inaccurate.

PART III: CONCLUSIONS

The public demands standardized achievement testing, and the uses of the scores increasingly involve high stakes. Elected representatives have responded to the public's demand for testing. The federal government has long maintained a national testing program, the National Assessment of Educational Progress, that measures student achievement over time. Standardized achievement testing programs in most states and

virtually all school districts provide information to help teachers assess student learning, plan better instruction, and inform policymakers and their constituencies. Their goals for student learning include appropriate curriculum, aligned instruction, and assessment based on multiple sources of valid information. But in most circumstances, the basis of assessment and accountability is a single source of information: the standardized achievement test. National education organizations have argued that test-based accountability is shortsighted, narrow, and inadequate. Given the frequent tendency for such scores to be inaccurate, we may in fact be doing more damage than good to our students. To continue using standardized achievement testing, we need to assure the public that our interpretation and each intended use are valid, not flawed or contaminated, as they often seem.

The many threats to the validity of standardized achievement tests this article has pointed out should concern us all. We need to evaluate these threats honestly and minimize or eliminate them. Without documentation or research that dismisses or qualifies such threats, it is hard to justify the public's longstanding confidence in standardized achievement test scores.

Perfection in test development and validation is unattainable (Phelps 2006). I agree. However, high validity standards are mandatory, particularly when test scores affect students' lives and their teachers' careers. If the messenger of student learning is so badly flawed, where is the truth in the message? The perils need more attention than they have received in the past.

References

- American Educational Research Association (AERA). 2000. "Position Statement of the American Educational Research Association Concerning High-stakes Testing in Pre-K-12 Education." *Educational Researcher* 29: 24-25.
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 1999. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Carroll, J. B. 1963. "A Model for School Learning." *Teachers College Record* 64: 723-733.
- Downing, S. M., and T. M. Haladyna, eds. 2006. *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Ferrara, S., and G. E. DeMauro. In press. "Standardized Assessment of Individual Achievement in K-12." In *Educational Measurement*, ed. R. L. Brennan, 4th ed. Westport, Conn.: American Council on Education/Praeger.
- Haladyna, T. M. 2004. "The Conditions of Assessment of Student Learning in Arizona: 2004." In *The Conditions of Pre-K-12 Education in Arizona: 2004*, ed. A. Molnar. Tempe, Ariz.: Education Policy Studies Laboratory at Arizona State University.

- Haladyna, T. M., and S. M. Downing. 2004. "Construct-irrelevant Variance in High-stakes Testing." *Educational Measurement: Issues and Practice* 23(1): 17-27.
- Haladyna, T. M., S. B. Nolen, and N. S. Haas. 1991. "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution." *Educational Researcher* 20: 2-7.
- Haladyna, T. M., and R. M. Olsen. 2006. Threats to validity in large-scale writing performance tests: What are these threats and what should be done about it? Paper presented at the annual CCSSO Large-Scale Assessment Conference, San Francisco, Calif.
- Haladyna, T. M., S. Osborn Popp, and M. Weiss. 2005. Nonresponse in large-scale assessment. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Kane, M. T. In press. "Validation." In *Educational Measurement*, ed. R. L. Brennan, 4th ed. Westport, Conn.: American Council on Education/Praeger.
- . 2006. "Content-related Validity Evidence." In *Handbook of Test Development*, eds. S. M. Downing and T. M. Haladyna, 131-154. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kolen, M. J., and R. L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices*. 2nd ed. New York: Springer-Verlag.
- Lohman, D. F. 1993. "Teaching and Testing to Develop Fluid Abilities." *Educational Researcher* 22: 12-23.
- McCallin, R. 2006. "Test Administration." In *Handbook of Test Development*, eds. S. M. Downing and T. M. Haladyna, 625-652. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Messick, S. 1984. "The Psychology of Educational Measurement." *Journal of Educational Measurement* 21: 215-237.
- . 1989. "Validity." In *Educational Measurement*, ed. R. L. Linn, 3rd ed., 13-104. New York: American Council on Education and Macmillan.
- . 1994. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Measurement: Issues and Practices* 23(2): 13-23.
- . 1995a. "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning." *American Psychologist* 50: 741-749.
- . 1995b. "Standards of Validity and the Validity of Standards in Performance Assessment." *Educational Measurement: Issues and Practice* 14(4): 5-8.
- Phelps, R. 2006. "Characteristics of an Effective Student Testing System." *Educational Horizons* 85(1).
- Sorenson, D. 2006. *2006 State Education Test Security Results*. Midvale, Utah: Caveon Test Security.
- Sternberg, R. J. 1998. "Abilities Are Forms of Developing Expertise." *Educational Researcher* 27(3): 11-20.
- Warren, J. R., K. N. Jenkins, and R. B. Kulick. 2006. "High School Exit Examination and State-level Completion and GED Raters, 1975 through 2002." *Educational Evaluation and Policy Analysis* 28(2): 131-152.

Tom Haladyna, Professor of Educational Psychology in the College of Teacher Education and Leadership at Arizona State University, has been an elementary school teacher, a test developer, a test researcher, and a teacher educator. He has published thirteen books, more than seventy articles and chapters, and hundreds of technical reports, validity studies, and conference papers. His primary interests are planning and validating testing programs, item development, and validity studies. He has worked in different capacities in more than fifty testing programs. He is co-editor, with Steve Downing, of the Handbook of Test Development (Lawrence Erlbaum Associates 2006).

Testing for Justice

by Edward G. Rozycki

There is one story that, over the past seventy years, increasing numbers of school people in the United States have come to tell. The story expresses widely shared aspirations and deeply felt concerns. Let us call this story the “Testing for Justice Rationale.” It goes something like this:

For schools to meet the needs of some children and not others is unfair. Justice therefore dictates that we meet the needs of all children. How do we determine those needs? By comparing what children can do with what they can learn to do. Any discrepancy between achievement and potential indicates a need. How do we determine such questions as achievement and potential? Through adequate testing.¹

That rationale, though it supports many well-intentioned attempts at upgrading American schools, is replete with questionable assumptions seldom examined after repeated tries at improving schooling practice have failed.

Educational testing has long been noted to affect the lives of not only students but educators themselves.² Thus, an understanding of testing and the assumptions on which it is based is indispensable to intelligent schooling practice. Tests can be critiqued not only for their technical efficiency but also on whether they are fair and whether the very process of testing is little more than an exercise of political power. The Testing for Justice Rationale burdens testing with determining not only need but, ultimately, justice as well.

Why Have Tests?

Modern schooling, which processes large numbers of students, seems inconceivable without testing. That is because it is so convenient for sorting students. It can stand in for a long and involved set of social interactions with master teachers—more typical of an apprenticeship

and more common before schools grew to present-day sizes and pursued a philosophy of productive efficiency.

Why do teachers give tests? For several reasons, among them: a) to support the authority of the teacher's judgment about acquired learnings, and b) to substitute for an infeasibly broad examination of student ability. This convenience is so important in the mass processing of today's schools that learnings not susceptible to easy examination, e.g., with paper and pencil, find it hard to gain status in a curriculum. Goodson comments,

For the groups and associations promoting themselves as school subjects, and irresistibly drawn to claiming academic status, a central criterion has been whether the subject's content could be tested by a written examination for an able clientele.³

In testing, however, we make many crucial assumptions about means, ends, and the causal connections among them. Achievement tests, for example, are not in and of themselves the point of instruction; otherwise, we would teach, not merely *to the tests*, but the very tests themselves. Nor is mere participation in course work thought sufficient to make testing unnecessary. Rather, the ends sought in achievement tests are certain important residues of the instructional process.

Calling something a test assumes a strong consensus on what its results indicate. But for constructs as vague and controversial as human abilities, upon which a judgment of educational need might be based, interesting things happen. On one hand, tests may stand in for controversial and pluralistic conceptions of human ability. Intelligence, for example, becomes what IQ tests measure. On the other hand, the concept of, say, intelligence itself becomes a focus of controversy.⁴

What Makes a Test a Test?

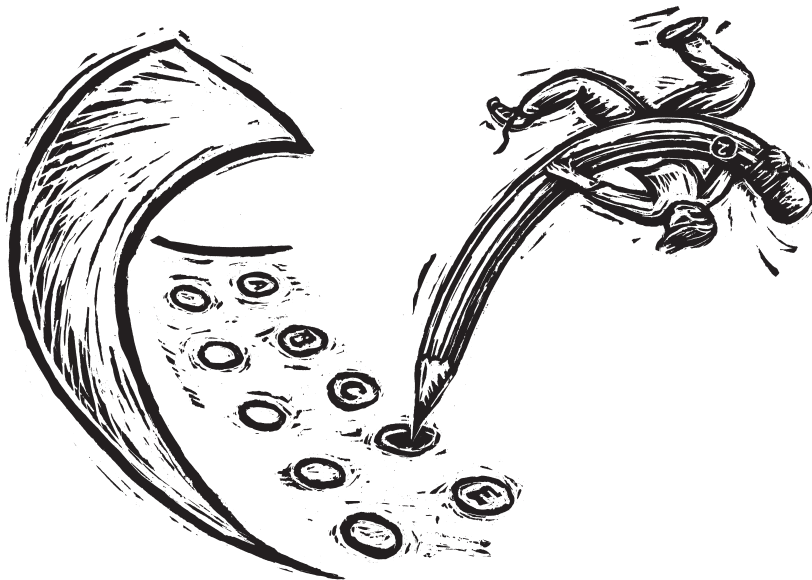
From the student point of view every test is a task.⁵ But not every task is a test, even if it looks like one. What conditions must a task satisfy to constitute a true test? It is a question of great practicality. State governments base school district funding on efficiency, itself determined by tests that state departments of education impose on the districts. But what will make the procedure anything more than a charade?

To avoid overlooking assumptions built into our conception of testing, let's substitute a different concept, *rank-task*, for tests. A rank-task is a type of activity for which some outcomes can be ranked: better, the same, or worse. Think of a rank-task as any procedure that assigns a number. It can be interpreted as a rank to compare that person to others involved with the procedure. Cinderella's prince, looking to fit the glass slipper, would be undertaking such a rank-task. Some feet are too small;

others, too large; only Cinderella's, just right. Trying to sort football players by the numbers on their jerseys is not a rank-task, though, because there is generally no significance to the comparison of any two numbers other than indicating a different wearer.

Tests are, at the minimum, rank-tasks. They can be performed with more or less skill. But the skill demonstrated may not be what we wish to measure. For instance, students take SAT-preparation courses to learn test-taking skills, not the information the tests are designed to measure. Often test-taking skills can be as critical to earning a good score as actual knowledge of the material the test covers. For some years, for example, the Princeton Review's basic test-taking materials and training have evidently raised SAT scores significantly.⁶ The SATs are intended to measure scholastic aptitude, but the effectiveness of the Princeton Review materials suggests that the SATs are also measuring something else—namely, the ability to take standardized tests of this type.

That observation illustrates the practical nature of our seemingly theoretical observations about testing. Among the readers of this article there certainly are individuals who did not receive a scholarship, or who were not accepted to the college or university of their choice, because of the scores they received on the SAT. And there is a fair chance that the reason for those scores was lack of, not scholastic aptitude, but certain test-taking skills.



Tests are also taken to be indicators. As such they must meet certain conditions. Any well-designed rank-task must be able to vary consistently upon reapplication, and the variation must be understood to make a difference. Test makers call that trait consistency, or internal validity. Tests must also indicate something other than themselves: that is called “external,” or construct, validity.⁷

Usually out of the hands of professional test makers is a fourth condition, trustworthiness: we must be able to believe the results were not manipulated for special purposes. That is usually a matter of test security, a matter with which many schools not infrequently deal in cavalier fashion.⁸

The important point, especially from the test taker’s point of view, is that every test is a task that can be performed, independently of such technical considerations as externality and trustworthiness, with greater or lesser skill. For example, a student may learn to take multiple-choice exams efficiently even if those exams test nothing recognizable as subject matter; yet a student who knows a great deal about some subject may falter at demonstrating that knowledge on the test prepared for it.⁹

If a rank-task is a test, then the goals of the testing control (determine) the kinds of test tasks we present to the student. Those tasks in turn control the knowledge the student will have to bring to support the test task. The connections between student knowledge and the test outcomes used to evaluate it are mediated by the task itself. Whether an increase in test scores indicates an increase in student knowledge or an increase in test-taking skills may depend on such mediation.

From Consensus, through Testing, to Justice

Let’s reiterate an important point: we, as interested parties, must agree upon some way of determining student knowledge independent of the test; otherwise the test becomes problematic. Lacking such consensus on the test, evaluations of potential or achievement are questionable. So then is the determination of need and consequently fairness.

Thus, in a very real way, problems of consensus are what bear ultimately, via testing, upon perceptions of fairness in schooling. We can lay the argument out as follows:

- a) Consensus, among interested parties, will affect which ideas of potential (e.g., native ability, capacity, competence) can be used for testing in the school.
- b) Consensus will affect which ideas of achievement (e.g., skills acquired or developed) can be used for testing in the school.

We then bring in the connections given by the Testing for Justice Rationale:

- c) The difference between potential and achievement measures need.
- d) The difference in treatment of need measures justice.

The most immediately practical version of this argument, which we will call the Status Quo Argument, is this:

There is a consensus in our community that Group A and Group B differ in potential. We observe that they differ in achievement. Since their achievement merely reflects their potential, there is no disparity in educational need. Therefore, our present treatment of Groups A and B, although they may look different, is not unjust.

The Status Quo Argument is theoretically sound, despite the fact that it has been pressed into the service of racism and class bias.¹⁰ The moral issue is how its supporting consensus arises. It is around such claims of consensus that many of the controversies about schooling cluster. (Consider, for example, the widely accepted assumption that so-called “gifted students” have no need for special educational treatments.)

Objectivity and Need

One assumption of much discussion about schooling practice is that testing offers an objective decision-making procedure that avoids problems of values and consensus. But is that so?

Test data seem so impartial, so objective. But what can numbers alone tell us? Imagine three groups of students, A, B, and C, who each receive a rank-task: Rank-Task 1, Rank-Task 2, and Rank-Task 3.

Suppose chart 1 gives us the following results—assuming the group means to be calculable.

Chart 1

	Rank-Task 1	Rank-Task 2	Rank-Task 3
Group 1	95	95	60
Group 2	50	50	60
Group 3	15	15	60

Even if we can also assume the significance of intergroup differences for each test and the absence of cheating, what are we to make of the differences in these scores? Are they any guide to practical decision-making?

It all depends. Our first question should be “What are these tests supposed to indicate?” Unless we believe they indicate something, they are still merely rank-tasks. And test results that are important to making equitable schooling decisions must deal with what Thomas Green has called *educationally relevant attributes*.¹¹ An attribute is educationally relevant in Green’s terms if it would be fair to distribute schooling benefits based on that attribute. If we believed it fair, for example, that males receive more diplomas than females just because they are males, then sex would be an educationally relevant attribute.

In America, unlike some other cultures, sex is by law not educationally relevant in public schools. Let us imagine a society so fixated on gender stereotypes that psychological distinctions override the physiological. To the extent that a female were seen as a “tomboy,” she would receive preference with “real men” over other females. “Girly men” would be devalued. In such a society, the test that decided who enjoys the privileges of gender prejudice would be called the “Degree of Masculinity Test.”

In chart 1, suppose Test 1 indicated something like “degree of masculinity” (DMT). If Test 2 indicated the percentage of high school graduates in the group, we in the United States would find that it indicated an unjust situation, because we reject gender as educationally relevant. But if instead Test 3 stood for the percentage of high school graduates, it would be taken, on the same assumption of the irrelevance of gender, to indicate equitable schooling practice.

More Educationally Relevant Attributes

Chart 2 (see next page) shows attributes in terms of which people might be grouped compared with different kinds of schooling benefits. In each block the words *just* or *unjust* indicate whether there is a general consensus in the United States that any schooling benefits distributed by the indicated kinds of attribute are considered just. Question marks indicate controversial practices.¹²

The chart indicates that in different situations an attribute may be educationally relevant, or it may not. Consider the case of sex-based grouping for varsity sports. Sex is generally not considered a relevant attribute so far as any educational benefit is concerned. Distributing high school diplomas based on sex, for example, is unjust. But participation in varsity sports is another matter. There is sometimes controversy about allowing women to play football, particularly in public high schools. Our chart indicates that with a question mark. (Imagine how chart 2 would look if it reflected common opinions in the United States circa 1800.)

Chart 2

ATTRIBUTE	BENEFITS						
	High school diplomas	Access to further schooling	Knowledge per se	Playing varsity sports	Nurturance	Special Programs	
SEX	unjust	unjust	unjust	?	unjust	just	
RACE	unjust	unjust	unjust	just	unjust	just	
HEIGHT	unjust	unjust	unjust	just	unjust	just	
ABILITY	just	just	just	?	unjust	just	
EFFORT	?	?	just	just	unjust	just	
CHOICE	just	just	just	just	?	just	
NEED	unjust	unjust	unjust	unjust	?	just	
WEALTH	?	unjust	just	?	unjust	just	
DISABILITY	just	just	?	just	unjust	just	
POTENTIAL	just	just	just	just	unjust	just	
ACHIEVEMENT	just	just	just	just	unjust	just	

Choice is an important and controversial attribute in our culture. It is not generally considered unjust if adults who decline to participate in certain programs, for example, fail to benefit from those programs. Lack of participation by children or mentally incompetent persons is often taken as a sign of immaturity or incompetence. Truancy is an example. Significantly, the lack of consequent benefits in truancy is still often argued as unjust, despite the insinuation that coercion may be justified. (This sense of injustice no doubt supports compulsory-schooling statutes.)

Other controversial practices suggested by the chart are:

- a. allowing students to play varsity sports on the basis of choice (interest) rather than ability (a long-established practice at Swarthmore College);
- b. social promotion—promoting students on effort rather than knowledge;
- c. providing nurturance, a scarce resource, on need rather than traditional practices of sharing per capita (i.e., “special education”);
- d. providing diplomas and sports participation based on wealth. (That is an important service of some kinds of private schooling.)

Needs and Consensus

Embedded in the Testing for Justice Rationale is an interesting equation:

$$(Ability) - (Achievement) = (Need)$$

Read this as “Ability minus achievement equals need” or “The measure of need is indicated by the difference between ability and achievement.” The equation often sorts students into three types: underachievers, normal achievers, and overachievers. Chart 3 shows several hypothetical scores for tests of ability and achievement. Using the equation given above, need is calculated. Based on need, students are typed as overachievers, normal achievers, and underachievers.

Chart 3

	Ability	Achievement	Need	Type
Group A	50	95	-45	Over-achiever
Group B	50	50	0	Normal achiever
Group C	50	15	35	Under-achiever

So it is argued that underachievers have greater educational needs—and the numbers make it seem quite objective.

Vague formulas like the one above, which can be discerned in the rationales offered, guide a surprising amount of daily school practice.¹³ They express not only accepted generalizations from practice but also conceptions of human nature. Their usefulness is not that they provide exact measures of important pedagogical constructs, but that they can so readily guide practice. But do they really identify needs? It depends on what we mean by needs.

In schooling, needs have long been treated as independent of consensus. But underlying much discussion of needs is the assumption that something should be desired. Someone who refers to a need is often urging action to address it while begging the crucial question of why we should address it.

We can distinguish between two conceptions of need: a conditional concept and an approval concept. A clear picture of the distinction can be obtained by comparing the following situations:

Situation 1: Johnny asks you to borrow a permanent marker. “I need it to write graffiti on the boys’ room wall,” he explains.

Situation 2: Mark tells you, “I need a permanent marker to complete my school art project.”

We would deny that Johnny needs a permanent marker but concede that Mark needs one. Why? Because we do not approve of graffiti, but we value art projects. If our values were different, our assessment of needs would be different.

The conditional concept of need says merely: some item X is necessary to bring about some other item Y. The permanent marker stands in this relation to covering the wall with graffiti as it does to doing the art project. In the conditional sense, both Johnny and Mark have needs, just as cars need fuel or terrorists need explosives. A conditional need indicates, at most, a lack. But lacks do not necessarily beg for remediation.

Talking about needs in schooling transforms an objective, take-it-or-leave-it conditional need into a need that elicits our support without careful consideration. The common technique is to show a lack of some kind and then to treat that lack as synonymous with an approval concept of need. A typical instance goes something like this:

Researchers working for one or another special-interest group announce with alarm that there is a great need to emphasize classical antiquity in the high school curriculum because 97 percent of five thousand high school seniors surveyed nationwide

could not identify Achilles, the Acropolis, Adonis, Aeneas, the *Aeneid*, and several dozen similar items.

If the research has been done properly, it does demonstrate that high school seniors lack knowledge of antiquity, but it does *not* demonstrate that we should do anything about it. That is an entirely different matter.

We are not disparaging needs slogans, merely reiterating the point that they assume and obscure issues of value and consensus. If, for example, people agreed on the value of “self-fulfillment” and what it means, then they would approve of what they believe is a causal or logical necessity to achieve self-fulfillment. An even more important consideration, though, is this: people may appear unmoved by appeals to needs not by heartlessness but by different values or beliefs in what is causally or logically related.

Examining the Rationale

Suppose candidates for school positions, teachers, principals, or superintendents were asked to comment on the Testing for Justice Rationale during their employment interviews. I would wager that were they to disavow or deny it, they would be denied employment (more likely, surreptitiously denied—moved to the bottom of the list—since we like to flatter ourselves that we are open to diversity in philosophy as well as race, religion, ethnicity, disability, or sexual preference—and lawsuits are expensive).

Too many schools, though, adopt such slogans as “All children can learn” or “We are dedicated to excellence”—and mandate that their staffs accept them. That leaves little wiggle room for those who find that the Testing for Justice Rationale presumes a near-blasphemous omnipotence.

Actually, if we analyze the Testing for Justice Rationale, we can see just where to distinguish issues of value versus issues of power. By doing so we may achieve consensus on important values without necessitating commitment to the possibly counterfactual optimism expressed in the Rationale. To revisit the Rationale:

For schools to meet the needs of some children and not others is unfair. Justice therefore dictates that we meet the needs of all children. How do we determine those needs? By comparing what children can do with what they can learn to do. Any discrepancy between achievement and potential indicates a need. How do we determine such questions as achievement and potential? Through adequate testing.

Is it really unfair for schools to meet the needs of some children and not others? Does the concept of readiness—so important to reading

teachers—not indicate our recognition that schools can meet the needs of the “ready” children better than others? And even if there is unfairness here, must it be schools that are responsible for addressing it? Does Justice dictate that? Or is it that other institutions in our society have foisted that off on the schools?

Ought we to accept responsibilities beyond the reasonable scope of our knowledge? In the long run, less blather about “determining potential” and more humility might enhance our professional repute to a greater extent than our posing as modern shamans for all things academic.

And if we are to accept such responsibilities, can we expect to be given reasonable resources to support our efforts? So far as funding is concerned, special education has been reneged on since its inception. Do we really expect a more generous flow from the public coffers in the future?

Testing is a side issue. Tests are constructed after most of the important issues—values, ethics, politics—that impinge upon schooling have been settled. That is why private and parochial schools are seldom consumed with the furor, the enthusiasm, and the dismay that testing brings to public education.

We may well continue to concern ourselves with the inequities we perceive in our society. We may well continue to pursue a dream of gleaming alabaster cities undimmed by human tears. If so, we might do better to look elsewhere than to public education to address our aspirations.

Notes

1. Cf. Nathan Glazer, “IQ on Trial,” *Commentary* (June 1981): 51–59, for a critique of *Larry P. v. Wilson Riles*, in which the Testing for Justice Rationale plays a prominent part.

2. Lynn Olson, “Tests Found Barring Thousands of Minority Teacher Candidates,” *Education Week* 8(12) (November 13, 1988): 1.

3. Ivor Goodson, *School Subjects and Curriculum Change* (Philadelphia: Falmer Press, 1987), 25.

4. For longstanding controversy on the nature of IQ, compare Arthur R. Jensen, “How Much Can We Boost I.Q. and Scholastic Achievement?” *Harvard Educational Review* (Winter 1969), with J. P. Guilford, *The Nature of Human Intelligence* (New York: McGraw-Hill, 1967). See also Howard Gardner, *Frames of Mind: The Theory of Multiple Intelligence* (New York: Basic Books, 1993).

5. Cf. Gilbert Sax, *Principles of Educational and Psychological Measurement and Evaluation*, 4th ed. (Belmont, Calif.: Wadsworth, 1997), 15: “A test is a task or series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes.”

6. Adam Robinson and John Katzman, “The Princeton Review,” in *Cracking the System: The SAT* (New York: Villard Books, 1986). See also the authors’ Web site at <<http://www.princetonreview.com/law/testprep/testprep.asp?TPRPAGE=17&TYPE=LSAT-PREPARE&adcode=23089>>.

7. For example, see Jum C. Nunnally, *Psychometric Theory* (New York: McGraw-Hill, 1967), 100-102.

8. See the NASSP's online article "High Stakes Cheating" at <<http://www.nasps.org/ContentLoad.do?contentId=151>>. Other subversive factors, such as cramming, are problematic. The Internet offers a surfeit of advice on how to cram effectively as well as admonitions that it will not work. See, for example, Eastern Illinois University's Learning Assistance Center's article "Final Exams and Cramming," at <<http://www.eiu.edu/~lrnasst/finals.htm>>.

9. See Linda M. Lance, "The Effects of Teaching Test-Taking Strategies to High School Special Education Students on Achievement Scores on the New Jersey High School Proficiency Assessment" (Diss., Widener University, 2004). Its abstract is at <<http://muse.widener.edu/~egrozyck/Dissertations/LanceR.html>>.

10. An inversion of this argument used to support giving more resources to those identified as "special" argues that such a differential is not unfair to "normal" children.

11. See Thomas F. Green, *Predicting the Behavior of the Educational System* (Syracuse, N.Y.: Syracuse University Press, 1980), 49-52.

12. These results are meant merely to be illustrative for the purposes of the exposition.

13. Other slogan formulas commonly encountered are *potential + learning = ability*, *ability + motivation = performance*, *achievement = sum(performances)* and *potential ≥ achievement*.

*Edward G. Rozycki, Ed.D., a contributing editor for **educational HORIZONS**, is an associate professor of education at Widener University, Widener, Pennsylvania.*

Publishing in educational HORIZONS®

educational HORIZONS requests that potential contributors or guest editors consider the topics in the adjoining space or suggest related themes or articles for upcoming issues of the journal.

Book reviewing:

Individuals interested in submitting book reviews (including more substantial book review essays that would review relevant scholarship on the topic) are encouraged to query by first-class letter or e-mail. Proposals, which can be independent of our issue themes, should specify recent book releases that will interest our readership of teachers and teacher educators.

For more information, contact:

educational HORIZONS®
P.O. Box 6626
Bloomington, IN 47407-6626
publications@pilambda.org

- Where Is Public Schooling Headed?
- Teacher Education: What Must Be Done?
- How Pertinent Is Education Research?
- The Great Grade Inflation, Its Causes, and Its Effects
- America's School Funding Fiasco: The Evils of Inequity
- Are Textbooks Obsolete?
- Silk Purses from Sow's Ears: Are the Broad Masses Truly Educable?
- The Classroom Impact of Multiple Intelligences Doctrine
- A Report Card for No Child Left Behind
- Are We All Constructivists Now?
- The Reading Wars: Where Are We Now?

Data-based Decision-making: Three State-level Educational Leadership Initiatives

by Van E. Cooley, Jianping Shen, Deborah S. Miller, Peter N. Winograd, John Mark Rainey, Wenhui Yuan, and Lisa Ryan

The accountability required by school-reform measures such as No Child Left Behind has placed increased emphasis on data analysis for appraising schools, administrators, teachers, and students. The focus of accountability-driven initiatives is developing policies and procedures that collectively influence the district, the school, and most important, the classroom (e.g., O’Day and Smith 1993; Shen and Ma 2006; Smith and O’Day 1990).

The Wallace Foundation has awarded grants focused on enhancing educational leadership in data-based decision-making. The foundation’s State Action for Education Leadership Project II (SAELP II) awarded grants to state departments of education in Ohio, New Mexico, and Michigan to develop school leaders’ data-based decision-making. This article describes the progress and potential of those initiatives.

The Ohio SAELP II Initiative: Training School-leadership Teams

Ohio’s SAELP II initiative trains school-leadership teams to access, interpret, and apply value-added information obtained from the statewide accountability system. The Columbus Public Schools (CPS), the Columbus Education Association, Ohio State University (OSU), and Battelle for Kids contributed to developing the program’s “train the trainers” model for elementary and middle school “All-School Improvement Teams” (ASIT). Two Regional Value-Added Specialists (RVAS) taught each leadership team—a principal, a building union representative, two other classroom teachers, and a parent—how to use value-added information

in redesigning preparation and professional-development programs. OSU's graduate-level classes for members of the CPS leadership teams allowed participants to choose from various training modes to address different learning styles and scheduling needs. Options included workshops, self-paced data e-school modules, and graduate-level OSU course work featuring face-to-face and online instruction.

The training developed data-based decision-making skills by teaching team members how to:

- use value-added information, along with other forms of data analysis, to improve student achievement with informed decisions;
- report student information by quintiles and different demographic groups;
- use value-added reporting to monitor year-to-year student progress, regardless of achievement level;
- focus professional-development efforts on areas of greatest need;
- create an ASIT network for sharing information and practices related to value-added reporting; and
- identify best practices for raising student achievement to replicate in other classrooms and schools.



CPS piloted the training model and obtained feedback from school personnel to ensure that the final training model met school-site needs. Despite several challenges that inhibited progress—mobility of CPS staff, scheduling difficulties, and lack of value-added data for CPS high schools—all CPS elementary and middle school teams received a full day of training in using value-added information, including opportunities to experiment with manipulating data. In addition, a core group of elementary teachers, including team members, took the OSU course on data analysis and value-added information to deepen understanding of the data. Trained school leaders have used the value-added growth model to analyze the educational needs of students who are not showing a year or more of growth. The analysis indicates that CPS students are having the greatest difficulty with assessments that include high-level, multi-part items and items requiring extended or short responses. Benchmarking measures that validate student progress have helped build and maintain teacher morale.

CPS educators should thus be better equipped to interpret and use value-added information when the value-added progress measure is added to the Ohio Accountability System in 2007–2008. In preparation for that change, Battelle for Kids (www.battelleforkids.org) and the Ohio Department of Education (www.ode.state.oh.us) are sponsoring a comprehensive skills-development program for a cadre of educators representing Ohio's twelve school-improvement regions. Battelle for Kids is offering a toolkit of interactive and print materials to complement the program. Beginning in 2005–2006, eighty Regional Value-Added Specialists made a two-year commitment to learn more about the uses and benefits of value-added analysis and to provide training for other educators.

In addition, Ohio is developing a new data-exchange system, D3A2 (www.d3a2.org), to support decision-making at the state, district, school, and classroom levels. A collaborative effort of data users and providers, the system is designed to provide timely and accurate data to all users—from the classroom teacher customizing instruction based on data to the researcher conducting longitudinal analysis of student academic growth.

The New Mexico SAELP II Initiative: Accountability Literacy

New Mexico faces particular challenges among students from culturally or economically diverse backgrounds. New Mexico ranks near the bottom on national measures of reading and math achievement (e.g., NAEP) as well as in poverty and student well-being. To address the challenges, in 2003 the state approved the Public School Reforms Act, which calls for key elements of systemic reform:

- challenging curricula;
- a three-tiered, performance-based teacher-licensure system;
- a more responsive governance structure;
- stronger accountability and assessment systems; and
- an aligned system of support for students, families, and schools.

New Mexico's SAELP II initiative focuses on ensuring that administrators and other educational leaders receive the resources necessary to use accountability data effectively. The New Mexico Department of Finance and Administration's Office of Education Accountability (OEA) are leading the SAELP II initiative. OEA has established a partnership with the Public Education Department, the Children's Cabinet (state agencies that deal with children and youth), the New Mexico Coalition of School Administrators, and six demonstration school districts.

During the first year of the grant, the partnership sought to answer three key questions:

1. What kinds of accountability data do principals, superintendents, and other educational leaders need?
2. What constraints do principals, superintendents, and other educational leaders encounter in obtaining and using accountability data effectively?
3. How can the SAELP II project help remove those constraints?

The answers to those questions were not surprising. First, principals, superintendents, and other leaders need data on student achievement and teacher effectiveness, as well as data that could show trends and could be used to improve instruction. Second, educational leaders face problems such as getting data on time and agreeing what data are needed. Third, educational leaders want the SAELP II initiative to ensure that

- data are gathered and disseminated effectively and used for appropriate and constructive purposes;
- leaders have the time, resources, and authority to make decisions based on the data;
- training and data-mining tools are developed; and
- improvements are made in how leaders are prepared in the university and supported through meaningful professional development.

The answers to the three key questions can be organized into New Mexico's Hierarchy of Educational Leaders' Data Needs (adapted from Maslow), as shown in the illustration.

The answers to the three key questions also contributed to the idea of "accountability literacy." The term refers to an educational leader's ability to understand the strengths and weaknesses of accountability data; to use that data to negotiate support for education in political, professional,

and community settings; and to improve students' lives by using data to argue effectively on their behalf. Information gathered through the SAELP II initiative reveals that many educational leaders in New Mexico have difficulty even obtaining access to important data, while others have struggled to develop accountability literacy on their own.

To address those problems, SAELP II has taken the following steps:

- developing data tools, including pivot tables and a data-based decision-making Web site, to help principals and superintendents analyze student achievement based on New Mexico's standards and benchmarks;
- creating a principal support network that provides professional development in accountability literacy to a cohort of principals from eight school districts;
- helping principals and superintendents effectively use New Mexico's new Student Teacher Accountability Reporting System (STARS) to improve student achievement;
- working with New Mexico's universities to improve recruitment, preparation, appraisal, and professional development for principals; and
- creating the New Mexico Children's Budget to track the funding allocated to youth-serving programs across all facets of state government.



More detail about each of those activities and other efforts associated with New Mexico's SAELP II initiative can be found at <www.nmsaelp2.org>.

The Michigan SAELP II Initiative: A Coalition Approach

Michigan's SAELP II initiative formed the Michigan Coalition of Educational Leadership, which seeks to improve educational leaders' data-based decision-making skills at several levels. The coalition includes the governor's office, the Michigan Department of Education, the major professional organizations, and three universities that prepare more than 50 percent of Michigan's educational administrators. The coalition has focused on four major tasks:

- developing demonstration sites for data-based decision-making;
- connecting the effective use of data with the Michigan Framework for School Improvement;
- infusing data-based decision-making into the professional development and endorsement of professional associations; and
- strengthening data-based decision-making instruction in educational-leadership programs at three universities.

Sixteen principals from four urban school districts with large populations of disadvantaged students participated in developing demonstration sites for data-based decision-making. The initial activity was a context analysis based on interviews with the principals. The principals responded to questions about their comfort in using data, barriers they had encountered in using data, and decisions they had made based on data. The interviews revealed that the principals struggled with time constraints, felt overwhelmed by the massive amounts of data, and lacked knowledge about ways to use data streams from multiple data sources to improve student achievement.

The principals then participated in two three-day retreats and five workshops. During the activities, the principals received the following types of training:

- an overview of balanced leadership (Marzano, Waters, and McNulty 2005) and "what works in schools" (Marzano 2003);
- professional development on data-analysis strategies for input, process, and output data;
- information on strategies for linking data to curriculum, instruction, and student achievement; and
- instruction on development of high-impact strategies to support the new Michigan Framework for School Improvement.

As part of a systematic data-mining process, the SAELP II initiative connected the Michigan Educational Assessment Program (MEAP), district-administered norm- and criterion-referenced tests, Standard and Poor's data, and other data sources. The principals began to understand the meaning behind data streams and data monitoring, the importance of benchmarking, and using multiple data sources to connect data with high-impact strategies in curriculum and instruction. In a working session, the principals met with their school-improvement teams to explore the uses of data for decision-making.

The SAELP II initiative is now applying lessons learned from the demonstration sites to implement projects involving the Michigan Framework for School Improvement, voluntary certification of administrative leaders, professional development and endorsement, and university-based educational leadership programs. The new Michigan Framework for School Improvement comprises five strands: teaching and learning, instructional leadership, personal and professional leadership, school and community relations, and data and information management. Each strand includes a data component. The SAELP II initiative complements statewide introduction of the Michigan School Improvement Framework by emphasizing the use of data to identify a school's strengths and weaknesses and to develop high-impact strategies for promoting student achievement.

In collaboration with the Michigan Department of Education, the SAELP II initiative is creating data-based decision-making modules for voluntary certification. In the early 1990s the Michigan legislature abolished administrative certification. Anticipating the passage of a new voluntary-certification bill, the Department of Professional Preparation Services has assembled a committee to help develop standards for certification and endorsement. Data-based decision-making will be an integral component of voluntary certification.

Recently the Michigan Department of Education and major professional associations received a grant to develop a professional-development program for principals for administrative endorsement. The Michigan Leadership Improvement Framework Endorsement (MI-LIFE) project is developing a leadership-training curriculum based on the new Michigan School Improvement Framework. The MI-LIFE endorsement program will include strong components on data-based decision-making. Marion Ginopolis, the director of the MI-LIFE project, asserts: "The integration of data analysis that the Wallace Foundation Grant Project [SAELP II] adds to each of the courses of the MI-LIFE leadership curriculum will elevate the MI-LIFE program from best practice to exemplary leadership." Because the legislature is passing a bill that allows professional associations to provide endorsements to school administrators, a data-based

decision-making component also is being developed for major professional organizations' endorsement programs.

Finally, SAELP II is enabling professors from Central Michigan University, Eastern Michigan University, and Western Michigan University to develop modules on data-based decision-making for their respective educational-leadership programs. Such programs, complementing the work of other SAELP II projects, will help achieve the goal of equipping all educational administrators in Michigan with effective data-based decision-making skills.

Coda

The SAELP II initiatives in Ohio, New Mexico, and Michigan engage all major state stakeholders in systematically improving student achievement through data-based decision-making. The three states' approaches provide a reservoir of information on promoting data-based decision-making. Data-driven leadership is not a fad: it will continue to influence decisions made by teachers, principals, boards of education, and other educational leaders. The lessons learned from the SAELP II initiatives of those three states are likely to affect many states, districts, and classrooms in the near future.

References

- Marzano, R. J. 2003. *What Works in Schools*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Marzano, R. J., T. Waters, and B. A. McNulty. 2005. *School Leadership That Works*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- O'Day, J. A., and M. S. Smith. 1993. "Systemic Change and Educational Opportunity." In *Designing Coherent Policy: Improving the System*, ed. S. H. Fuhrman, 250-312. San Francisco: Jossey-Bass.
- Shen, J., and X. Ma. 2006. "Does Systemic Change Work? Curricular and Instructional Practice in the Context of Systemic Change." *Leadership and Policy in Schools* 5(3): 231-256.
- Smith, M. S., and J. O'Day. 1990. "Systemic School Reform." In *Politics of Curriculum and Testing*, ed. S. H. Fuhrman, 233-267. London: Falmer.

Van E. Cooley and Jianping Shen are professors of educational leadership at Western Michigan University; Deborah S. Miller is the Ohio SAELP Project Manager, Ohio Department of Education; Dr. Peter N. Winograd is the Director of Office of Education Accountability, New Mexico Department of Finance and Administration; John Mark Rainey, Wenbui Yuan, and Lisa Ryan are doctoral students in educational leadership at Western Michigan University.

Challenges in Data-based Decision-making: Voices from Principals

by Patricia L. Reeves and Walter L. Burt

The Case for the Principal as Shaper of Data-based Decision-making

The past twenty-four months of educational leadership literature reveals a steady stream of prescriptions for how a principal should shape the focus of a school: by raising student achievement through shared leadership, data-based decision-making, and unwavering attention to the employment of best practice in curriculum, instruction, and assessment (Marzano, Waters, and McNulty 2005). If we read between the lines, the importance of systematic collection, analysis, and interpretation of multiple data sources and types becomes much clearer. Principals in the information age need to be information driven, committed to shared leadership and relentless about continuous improvement. They must reshape the processes, norms, and behaviors of teaching and learning (Glickman, Gordon, and Ross-Gordon 2004) around aggregating and interpreting shared information, i.e., data (Picciano 2006).

The theoretical and the research literature are both increasingly consistent and clear in redefining the school principal's archetypal building-management role (Reeves 2004). Implementing the new definitions of school-leadership preparation and practice standards will help create schools that continuously learn and adapt to student needs and improve student outcomes (Engler 2004). The cornerstones of those standards—assumptions of moral purpose, transformational processes, inclusion and diversity, and a culture of safety and success for all students—and the roles they shape for school leaders involve creating the conditions for continuous learning and change (Fullan 2001); thus, each bears close monitoring through multiple information or data sources.

Ultimately, today's school principals are charged with two challenges: 1) breaking through entrenched, loosely coupled systems (Cusick

1992) characterized by work isolation, uncertain technologies, and top-down decisions; and 2) forging new dynamic and flexible systems that intensively engage new precision technologies, data-informed processes, and collaborative decisions (Lambert 2003). Many new principals assume responsibility for schools already in various stages of federally mandated sanctions caused by achievement lags and gaps. Those principals and the schools they lead have no time for slow, incremental change. Moreover, they cannot afford to continue supporting the same teaching and learning processes. Their schools need new high-yield, strategic decisions based on deep understanding of the school context, student needs, and student performance profile to help ever-more diverse and more socioeconomically challenged student populations. A critical pathway to such in-depth understanding is data collection, analysis, and interpretation (Schmoker 1996).

Challenges in Data-based Decision-making

Effective shared decision-making thus requires knowledge, skills, and dispositions conducive to systematic gathering, analysis, and interpretation of relevant data. District leaders must understand the direction and training needed to support such leadership. One good way of learning what principals need is to go to ground level with them as they learn about and attempt to employ data-based decision-making strategies in their schools. The Michigan State Action for Educational Leadership Project II (SAELP II)



is doing just that, as part of a multi-state initiative funded by the Wallace Foundation with a focus on connecting educational leadership at the state and district levels. By working with sixteen principals from four urban school districts over a three-year period, the project has afforded its research team an in-depth look at the actual experiences of principals attempting to reshape the decision-making processes of their schools.

After a year of working with these sixteen principals, the Michigan SAELP II researchers are beginning to isolate common themes that relate to the conditions principals face in their school districts as they implement data-based decision-making. Through one-on-one interviews, the researchers collected and analyzed principals' descriptions and depictions of their status at the onset of the project. This analysis yielded four major themes: (a) teacher and principal knowledge; (b) teacher and student issues; (c) data overflow and other barriers; and (d) time to receive and analyze data.

Teacher and Principal Knowledge

Principals expressed concern about their own lack of training and understanding of how to use data. Several admitted having fears about mathematics and data analysis. They expressed concern that their university administrator-preparation programs failed to prepare them to analyze data. Principals reported little common understanding with teachers and other district administrators regarding what data are important and what the data mean. One summarized the frustration with using data:

I'm not sure how data driven I am sometimes. You know, I am looking at the data, making decisions based on the data, but sometimes I do not know if we are looking at that correctly.

Principals reported that training is critical to enhancing teachers' understanding of data. One principal expressed concern about the lack of professional development before the start of school and prior to the beginning of the second semester and also stated the need for professional development geared to individual schools instead of districts.

Another expressed frustration about the lack of teacher and administrator knowledge in data-based decision-making:

Our teachers need to be trained. Our principals need to be trained. I know I'm not telling you anything you don't already know. . . . [I]n our principal meetings . . . we are talking about data and adjusting instruction. . . . I look around and I know there are people in the room who buy into it [data analysis], but they are not quite sure how to do it.

The principal continued: “We graduate from college and we don’t know everything just because our title says principal, assistant principal, or whatever.” The overall message is that principals and teachers must involve themselves in sustained, job-embedded professional development geared to understanding and using data to make effective decisions.

Teacher and Student Issues

Principals identified several issues that pose challenges in using data under the broad categories of teacher and student perceptions and attitudes: e.g., “Some students do not take testing seriously,” and “[The] use of data is not necessarily part of teacher training, and one of the roadblocks is the acceptance of the time that we used to collect data—is [this] time well used?” They noted teacher frustration with how many students get the same results after taking tests multiple times. One principal discussed the need for accurate analysis of multiple tests to provide information that could be used for student improvement. Various tests generate different information, but understanding how to disaggregate and effectively use various test results is a challenge.

Coupled with the perception of being overburdened with testing is the problem of teacher and student attitudes regarding testing’s importance. One principal captured teacher attitudes in the following statement.

They [teachers] think that we are just testing students to death. Everybody is just exhausted from testing. You really don’t have valid data because the kids just go through and they begin to mark whatever.

That principal recognized that teachers use student attitudes as an excuse, then project their own attitudes on the students—and thus further exacerbate the problem of seeing value in and utilizing test results effectively. Additionally, principals cited the lack of coordination between tests. For instance, a district might administer several tests (Iowa Test of Basic Skills, Metropolitan Achievement Test, state-mandated tests, and other screening, diagnostic, and criterion tests) with no mechanism to organize, plot, and analyze the data from the various tests so that teachers can use the information to impact student achievement. There is thus no easy way for teachers and students to know what the test data mean.

Data Overflow and Other Barriers

Principals identified several barriers to data-based decision-making, including excessive raw data, inadequate technology to use data, coordination, and data warehousing. They felt that the amount of data was overwhelming—“[J]ust too much data, and sometimes it is really hard to

choose which data is reliable for what your intended purposes are.” One principal recommended streamlining the data. A second stated that there were no connections between the various assessment instruments used in their buildings, and that this made it difficult to analyze and use the data with teachers. A third principal suggested that the amount of data collected was a “big hindrance” to his faculty and staff: “With all of the data collected, what pieces do you take out and use?”

Disaggregating to examine subgroup performance and breaking down data to analyze individual student performance were also identified as challenges. Putting the data together for a complete picture of students is important but difficult. One frustrated principal who wants “relevant” data complained, “We have so many pieces—I know that each piece has its own job, but it’s overwhelming . . . to get it all and make sense.” The principal, suggesting that the data often produced mixed messages for faculty, emphasized the need for common meanings and processes when using data to improve student achievement.

Principals also identified data warehousing and technology as barriers to data-based decision-making. One principal recounted difficulty in preparing data for teachers due to computer and printer malfunctions. The technology can lock up because of the amount of data. Related issues were (1) accessing data in a usable form; (2) disaggregating data to a point where it has meaning for teachers; and (3) receiving data in a time frame that makes it truly useful for intervening with student learning.

Time to Receive and Analyze Data

Principals identified time as a major barrier to providing leadership for utilizing data. Teachers do not have time to analyze data or to collaborate with one another regarding the meaning and use of data. Lamenting the lack of time for adequately monitoring teacher progress, as well as their own, in using data, principals contend that lack of time also influences teacher morale. Teachers feel stressed and unable to reach their goals. The following responses aptly represent such sentiments:

There is not sufficient time for teachers to meet and analyze the data. Teachers are busy and often do not want to do more than what they are contractually required to do.

There is not enough time for collaboration. When teachers get the data, they do not have the time to “mull” over it, talk about strategies, and think about how they can teach differently and share what they have done.

Following testing, getting the data back in timely fashion was, again, identified as a challenge. Educators forced to write their school-

improvement plans based on incomplete data, up to a year old, are always a year behind in their ability to truly use data to adapt teaching and learning processes.

Implications for District Support and Response

Based on the results of the principal interviews, major issues surrounding principals' experience with data-based decision-making obviously exist. Principals see little progress in connecting data use with classroom instructional decisions. Many principals, uncomfortable with data collection, analysis, and interpretation themselves, suggest that their teachers possess even less understanding of and appreciation for using data in decisions about classroom and school processes. They conclude that significant, targeted professional development on data collection and use is needed for both teachers and administrators. In addition, principals clearly need help in developing data-based strategies for monitoring teacher and student progress and fostering changes in attitudes, assumptions, and culture.

Principals recognize that using data effectively in today's schools suggests completely rethinking teachers' and students' roles in learning and decision-making. They perceive, however, that both students and teachers suffer a low sense of efficacy regarding assessment and shaping decisions with data. Principals agree that teachers must become problem solvers, but they also need the skills and time to do so. To help teachers collect and analyze data for every student, adapt instruction accordingly, and chart continuous improvement progress, their workday and responsibilities need restructuring. To match shared leadership with shared responsibility, any restructuring of the current teacher role must be accompanied by greater teacher efficacy and accountability. The issues run much deeper than merely collecting and analyzing data.

Finally, data overflow and disconnect must be addressed through district data collection, warehousing, and reporting systems. Most principals interviewed mentioned that problem and their frustration with the absence of reliable means to access, manipulate, and interpret data efficiently and promptly. Although fully acknowledging the need to utilize data from multiple sources systematically, they also emphasize that this is the area in which they need the most help from their districts; they simply lack the tools, expertise, and time to function at the current "every man for himself" level in accessing and utilizing data for building-level decisions. They need data organized in a way that tells the story of each student's achievement and the achievement profile of the entire school over multiple years. They also need coordinated district leadership focus and facilitation to utilize data-collection and analysis tools effectively in their school decision-making processes.

Substantive changes require addressing the conditions the data gathered in this first year of Michigan SAELP II project reveal. Although the conditions must change at the building level, intentional, systemic support at the district and state levels is necessary or principals will fight an ever-steeper uphill battle. Discussions with principals revealed animosity, cynicism, mistrust, and a general lack of confidence at both the district and state policy levels. As Salpeter (2004) observed, school improvement is influenced by many factors; consequently, systemic change in both the conditions for and the support of principals' new role as instructional leader, change agent, and data-based decision-maker will be critical for success.

References

- Cusick, P.A. 1992. *The Educational System: Its Nature and Logic*. New York: McGraw-Hill.
- Engler, C. M. 2004. *The ISLLC Standards in Action: A Principal's Handbook*. New York: Eye on Education.
- Fullan, M. 2001. *Leading in a Culture of Change*. San Francisco, Calif.: Jossey-Bass.
- Glickman, C. D., S. P. Gordon, and J. M. Ross-Gordon. 2004. *Supervision and Instructional Leadership: A Developmental Approach*. 6th ed. Boston: Pearson-Allyn Bacon Publishers.
- Lambert, L. 2003. *Leadership Capacity for Lasting School Improvement*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Marzano, R. J., T. Waters, and B. A. McNulty. 2005. *School Leadership That Works: From Research to Results*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Picciano, A. G. 2006. *Data-driven Decision-making for Effective School Leadership*. Upper Saddle River, N.J.: Pearson-Merrill Prentice Hall.
- Reeves, Douglas B. 2004. *Assessing Educational Leaders: Evaluating Performance for Improved Individual and Organizational Results*. Thousand Oaks, Calif.: Corwin Press.
- Salpeter, J.. 2004. "Data: Mining with a Mission." *Technology & Learning* 24(8): 30-32, 34, 36.
- Schmoker, M. 1996. *Results: The Key to Continuous School Improvement*. Alexandria, Va.: Association for Supervision and Curriculum Development.

Patricia L. Reeves and Walter L. Burt, both former school superintendents, are now faculty members of educational leadership at Western Michigan University.

CALL FOR PRESENTATIONS

TWO POWERFUL DESIGNS—ONE CONFERENCE

Strengths-based Education (The Gallup Organization)

Critical Friends Groups (The National School Reform Faculty)

WHO SHOULD ATTEND?

Teachers • Administrators • Teacher educators
Education students • Staff-development professionals

This conference will explore tools for achieving student engagement in higher education and K-12, and the structure and training that are prerequisites for applying those tools successfully. The National School Reform Faculty's Critical Friends Groups design provides the necessary framework and The Gallup Organization's strengths-based education program provides the necessary tools.

Workshop proposals are invited on the four focus issues below. All presentations, including general sessions and workshops by the National School Reform Faculty and The Gallup Organization, are scheduled for Richmond, Virginia, on Friday, July 27, or Saturday, July 28.

TO SUBMIT A PROPOSAL

Visit www.pilambda.org and click on Richmond 2007 or call 800-487-3411.
Proposal deadline: January 8, 2007.

FOCUS ISSUES

Building peer support networks

Eliminating teacher isolation and drawing on colleagues to improve teaching practice are the essence of professional learning communities in K-12, and increasingly in higher education as well. Workshops on this focus issue will highlight successful programs in which educators are inspired to come together collaboratively to examine student work, help each other improve teaching practices, and give each other feedback that is challenging without being threatening.

Legislative/administrative advocacy

Non-educators dominate the ranks of legislatures, tax authorities, and regulatory bodies at all levels of government—even local school boards. Teacher-leaders have learned to identify the important issues and present the educators' point of view to those non-educators. Workshops on this focus issue may address effective approaches to specific advocacy issues, e.g., testing, standards, curriculum, and performance measures. Workshops may also highlight general techniques and strategies for identifying key education policy makers, understanding education issues from their perspectives, gaining access, communicating a point of view clearly and effectively, and maintaining a constructive, ongoing relationship.

Increasing K-12 and higher-education student engagement

Increased student learning and teacher fulfillment are now understood to be results of increased student engagement. Workshops on this focus issue will highlight approaches to increasing student enthusiasm and focus. Approaches might be based on what is known about, for example, parent engagement, brain function, gender differences, classroom management, student diversity, mainstreaming, positive psychology, or strengths-based education. Workshops on new teacher support and mentoring are encouraged.

Senior interests

Retirees have the time to do things that working educators simply don't have time to do. Conversely, having a place of employment allows working educators to take for granted things that may not be available to retired educators, e.g., computer training, routine use of the Internet, insurance and investment management, online travel management, and e-mail. Some workshops on this focus issue will highlight opportunities for the constructive, creative use of time; others will highlight skills and knowledge that are not easily available outside the workplace.



**ENERGY
through
SYNERGY**

**Powerful Designs
for Leadership**