

Strategies for High School Accountability: A Response to Marion

Linda Mabry, Ph.D., Washington State University Vancouver

Reidy Interactive Lecture Series (RILS)

October 2005, Nashua, NH

co-sponsored by

National Center for the Improvement of Educational Assessment and West Ed

I apologize in advance for any subversiveness in my remarks or in the questions I'll raise. I should also begin with an acknowledgement that I agree with much of what Scott Marion (2005) has stated about high school accountability and with the implications of what he has said, but there are also some deep divides. My comments will emphasize those, only partly as a kind of devil's advocacy intended to engage thinking and discussion. I've organized my comments around three topics.

Goals and Impacts

The first of these is *goals and impacts*, with which I am raising for consideration whether the strategies we are using to reach our goals for accountability are getting us there, are obstructing the attainment of our goals, or are counterproductive. My point of departure is a statement made by Brian Gong (at RILS, 2005) that it's "not just technical issues; there's a human side" in educational measurement.

As a former third-grade teacher in a very poor school in Houston, Texas, my concern is just that: the impact of assessment and accountability policy on students and teachers. I am concerned when policy requires the impossible and depends on indicators which may be as likely to misrepresent as to represent educational quality. I'm concerned when Scott lists:

- all that *must* be done by schools, for example, that they must be held accountable for the post-secondary transitions of their graduates (Marion, 2005, slide 24) – when career choices actually depend on more than educational experiences and when few schools have the resources to track the lives of their graduates
- all that *must* be done by districts, for example, that they must provide assessment leadership and promote assessment literacy in schools (Marion, 2005, slide 28) – when just a few years ago, one of my colleagues, Mike Trevisan (2000), conducted a needs assessment in my home state of Washington to see what training was needed by district assessment officers and discovered that few districts had anyone very knowledgeable about assessment

- and all that *must* be done by states, for example, that they must ensure that district and school accountability systems are valid (Marion, 2005, slide 33) – when I can't remember the last time I saw a validation study I considered to be based on full and sufficient evidence.

So, the questions I raise for consideration are these:

Are we asking the impossible? Are we looking for a perfect world and ignoring the real world?

Are assessment and accountability, whatever their goals, *functioning* to punish the vulnerable? Are we complicit?

Are assessment or accountability solutions to *sociological* issues which are compounding *educational* problems?

What kind of educational accountability would be feasible? achievable? reasonable? – a question brings me to my next topic:

Who should be accountable for what?

Here, my point of departure is Scott's statement: "If we trusted course grades, we would not need any additional student-level high school accountability" (Marion, 2005, slide 38). But we don't. We haven't for 150 years, since the times of Horace Mann and Edward Thorndike. That is the reason we have standardized tests and test-driven accountability today. Just this morning, Kevin Stansbury (at RILS, 2005) noted problems with grade inflation in Louisiana.

Mark Musick (2005) pointed out the irony, perhaps hypocrisy, of Bill Gates (2005) calling high schools "obsolete" at a time when his own company, Microsoft, is being described in the media as "dysfunctional." But is measurement any better? The history of educational measurement has sometimes been described as a history of unintended consequences. Measurement folk (and plenty of others) are saying schools aren't good enough – when the *tests* aren't good enough. And the tests are not perfectible. We cannot eradicate measurement error; we can only estimate it using methods I consider highly suspect. *Accountability systems* aren't good enough – with their overreliance on a few, flawed indicators. But the accountability systems are not held accountable.

I am not claiming that schools do not need to improve, but by what method? The literature of educational reform has repeatedly shown that top-down approaches don't work very well, and systemic approaches have a checkered record. Still, that is what we have with No Child Left Behind (2001). A recent analysis of the policy development for Reading First, included in NCLB as Part B of Title I, showed that both professional education organizations and professional measurement organizations were peripheral to the federal policy development process (Miskel & Song, 2004). Perhaps unsurprisingly, NCLB is an unrealistic policy from either educational or measurement perspectives. But the policy nevertheless includes Title I money as a *bribe* and a whole

series of negative consequences to schools as *threats* – which puts the schools between the proverbial rock and hard place.

All this has occurred in spite of the fact that GPA (grade-point averages), based on teacher grades, are – as Laurie Weiss said this morning – data "that colleges find very useful" (2005). To this, Stan Rabinowitz added that, for many colleges, "SATs don't mean a thing" (at RILS, 2005). That is, we don't trust teacher grades although we know that grades are better indicators of college performance than college entrance tests are.

That point brings us back to Scott's statement – "If we trusted course grades, we would not need any additional student-level high school accountability" – and raises the questions:

Do we need accountability? Do we need this type of accountability?

Or is accountability, as Lee Cronbach (1988) said, a sign of a pathology in the system? Is what we are doing sick? Are we making the schools sicker?

As a researcher and program evaluator, I have interviewed teachers in a dozen states. I have never found one opposed to accountability or to being held accountable – or one who considered current accountability systems reasonable. Rather, teachers realize they are being held accountable for things over which they have no control. It has come to this: For the past seven years, when interviewees have occasionally shown themselves to be conspiracy theorists by describing testing or accountability as the means to undermine public schooling, they have done so privately, *sotto voce*. But, in my graduate assessment course this summer, our most prominent local district superintendent, a former state assessment director, described NCLB as a conspiracy to abandon public education – publicly and to people he himself might hire. So, I ask:

Can we do accountability better?

Would it be better to give something higher priority than accountability? For example, what if the resources now devoted to testing and test preparation were redirected to professional development, teacher training, school reorganization, or teacher salaries – paying enough to get and to keep the best and brightest?

Could we get more bang for the buck? Would that require us to trust teachers? Could we prepare and hire and retain teachers we could trust?

Representation

My third topic is representation. I certainly agree with Scott that multiple measures are needed in accountability systems. In the field of program evaluation, the minimum expectation for competent practice involves the collection and interpretation of relevant and sufficient evidence. Test scores and other numerical indicators are not enough. In evaluation, we know that educational quality can and should be determined on better evidentiary bases than we see in current educational accountability systems.

But with this topic, I'm speaking not as an evaluator but as a postmodernist. I'm not engaging in a philosophical exercise but, rather, applying postmodern skepticism and doubt to:

- (1) how assessment and accountability systems represent students and schools and
- (2) the hegemonic consequences of doing so.

Here, my main point is that *all representation involves misrepresentation*.

All test scores include measurement error. We don't even know how much. Scoring errors on state tests are reported every year. Every scoring corporation has committed errors on state tests, putting the states in a bind. There are troubling inconsistencies between different measurement results, which suggest that some measurement-based representations are in error. For example, in Oregon this year, there were a number of schools which the state considered satisfactory but which were considered unsatisfactory based on federal annual yearly progress (AYP) reviews (Carter, 2005; Hammond, 2005). Somebody misrepresented Oregon schools.

In my home state of Washington, we have had the highest SAT scores for two consecutive years among those states in which more than half of high school students take the SAT (Wood, 2005), yet only about a third of these same high school students are scored as proficient on the state's standards-based test, the Washington Assessment of Student Learning (WASL).¹ Does this not suggest measurement error somewhere? Does this not suggest that someone is misrepresenting our kids? In 2008, when passing the WASL will be required to get a high school diploma, our kids will pay for that error.

Another case in point, this time at the local level: I'm in the third year of an evaluation of high schools in Vancouver School District in southwest Washington state. One of the schools is Vancouver School for the Arts and Academics, which enjoys far and away the highest WASL scores in the district. Yet, in the current accountability climate, the district is running so scared that, two years ago, VSAA had to decrease the amount of time spent on the arts, its *raison d'être*, and increase the amount of time spent on academics – despite its high scores in academic subjects (Prince, 2003). Down the road is Fort Vancouver High School, where the district has concentrated its resources for English Language Learners and enrolled all high school ELL students in the district. Anybody want to guess which high school "needs improvement" according to AYP? Then there's Lewis and Clark High School, an alternative school, the type we ruefully chuckled earlier during this conference (RILS, 2005). Lewis and Clark has a 40% drop-out rate. Last year, some of its WASL scores were in single digits; even Fort Vancouver's scores were five times the scores at Lewis and Clark.

¹ See state websites: www.ospi.wednet.edu/ and www.k12.wa.us. The WASL is a conjunctive test. Taken separately, about two-thirds of Washington high school students pass the WASL reading and math tests.

Which is the better school? These indicators are not enough to tell; in fact, they could lead to misjudgments. Lewis and Clark is actually holding students accountable in the ways Scott suggested schools should do (see Marion, 2005, slide 18). Do you also need to know that transience of the student population at Lewis and Clark actually exceeds the enrollment?² Should – could – that be factored into for determining whether the school is performing satisfactorily? Would it help to know that graduate students in my assessment courses often make site visits to Lewis and Clark³ and that my students are routinely inspired – even awed – by the quality of the education and assessment they see at Lewis and Clark? that they often say they hope to be hired there or some place like it, where they might also make a difference? Lewis and Clark has actually figured out how to make assessment work *for* kids, even kids whom assessment has beaten down. The obverse to their 40% drop-out rate is a 60% resurrection rate. Still, the principal has not been able to sleep through the night for two years, worrying about accountability and whether his school may be closed, his staff lost, and his students abandoned.

While it is interesting philosophically to muse about whether educational measurement is a modernist enterprise, perhaps a last gasp of modernism, in a postmodern world, it is important to consider the questions which postmodernism raises for educational accountability:

Has testing become, as Foucault called it,

surveillance that makes it possible to qualify, to classify, and to punish. . . . [I]t manifests the subjection of those who are perceived as objects and the objectification of those who are subjected (cited in Rabinow, 1984, p. 197)? (see Marion, 2005, slide 46)

Are surveillance and objectification the implications of the "better data systems" Scott described?

Is testing a way to reproduce the social *status quo*, contrary to public education's aim to redress privilege through opportunity and meritocracy?

It concerns me that the tenor of our discussions suggests that we are, that measurement is, adversaries of high schools. We have talked about what is wrong with "them" and about what "we" should make "them" do. What I ask is this:

Can we make assessment and accountability work *for* poor kids, the teachers committed to teaching them, and their schools?

² In addition to student transience as it is usually understood, some Lewis and Clark students enroll, drop out, and re-enroll.

³ In addition, the principal and teachers at Lewis and Clark High School often take the time to come to my classes to share their experiences and insights. I again offer them my gratitude and admiration.

References

- Carter, S. (2005, August 12). State vs. federal ratings. *Oregonian*, p. A18.
- Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Gates, B. (2005, March 3). Fixing our obsolete high schools. *Oregonian*, p. D7.
- Hammond, B. (2005, August 12). More schools pass tests. *Oregonian*, pp. A1, A18.
- Marion, S. & Gong, B. (2005, October). Accountability at the high school level: Emerging recommendations for a multi-level set of issues. Paper presented to the annual Reidy Interactive Lecture Series, Nashua, NH.
- Miskel, C. & Song, M. (2004). Passing *Reading First*: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, 25 (2), 89-109.
- Musick, M. (2005, October). Building the foundation: Establishing our goals for high schools. Paper presented to the annual Reidy Interactive Lecture Series, Nashua, NH.
- No Child Left Behind Act (2001). Public Law No. 107-110. 107th Congress, 110 Congressional Record 1425, 115 Stat. (2002).
- Prince, A. M. (2003, August 15). VSAA panel presents compromise schedule. *Columbian*, pp. C1, C7.
- Rabinow, P. (Ed.) (1984). *The Foucault reader*. NY: Pantheon.
- Rabinowitz, S. (2005, October). Assessment approaches: Which options are best for what purposes? [discussant for paper presented by L. Wise]. Paper presented to the annual Reidy Interactive Lecture Series, Nashua, NH.
- Trevisan, M. S. (2000). Priority training needs of school district assessment coordinators in Washington state. *Studies in Educational Evaluation*, 26 (3), 231-244.
- Wise, L. (2005, October). Assessment approaches: Which options are best for what purposes? Paper presented to the annual Reidy Interactive Lecture Series, Nashua, NH.
- Wood, S. O. (2005, September 1). Washington outpaces Oregon on SAT – again. *Oregonian*, pp. B1, B3.