

## **Chapter 4**

### **Aiming at Learning: Assessment as the Critical Link**

**Mari Pearlman**

#### **Abstract**

Inequities in educational opportunity in the United States and variability in individual states' standards for student achievement have been amply demonstrated by the intensive focus on standardized test results over the past few years. The call for higher educational attainment for US students has intensified even as test results indicate persistent performance gaps. In order to effect real changes in the achievement of students in US public schools, particularly the learning achievement of children in poverty, we need to think differently about assessment, and recognize its rightfully central place in the education enterprise. To do this, however, we need also to recognize that we can make no progress without some common set of assumptions and benchmarks by which we gauge such progress. What we need is an assessment that is worth teaching to, one that allows for meaningful comparisons across states, and one that is connected with instruction in fundamental ways. One way to accomplish these goals would be to use the National Assessment of Educational Progress (NAEP) to create benchmark assessments at 4th, 8th, and 12<sup>th</sup> grade levels to assess the achievement of all students in US public schools against common standards for learning and performance.

**Keywords** Accountability · Assessment · Common standards · NAEP · NCLB · Performance benchmarks

No part of the public education enterprise has received more attention and engendered more rhetoric over the past 8 years than student assessment. The pressures brought to bear on educators through the enforced implementation of high-stakes annual testing of children and publication of the results have suggested that testing alone can change educational outcomes for children, a position somewhat akin to a belief that a really good thermometer can cure a fever. We have an opportunity in the next 2 or 3 years, with a new administration in Washington responsible for the reauthorization of the Elementary and Secondary Education Act, home of the No Child Left Behind legislation and attendant education policy, to focus on what really matters, which is learning. In order to effect real changes in the achievement of students in US public schools, particularly the learning status of children in poverty, we need to think differently about assessment, and recognize its rightfully central place in the education enterprise.

To do this, however, we need also to recognize that we can make no progress without a common set of assumptions and benchmarks with which to gauge such progress. We need an assessment that is worth teaching to, one that allows for meaningful comparisons across states, and one that is connected with instruction in fundamental ways. One way to accomplish these goals would be to use the National Assessment of Educational Progress (NAEP) to create benchmark assessments at 4th, 8th, and 12th grade levels to assess the achievement of all

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

students in US public schools against common standards for learning and performance. A number of challenges must be met to implement this idea: NAEP assessments would need to be designed for the reporting of individual student results, the NAEP content frameworks and performance standards would need intensive review and revision, and a great deal of conversation across state boundaries about the benefits of such an approach would need to occur.

In order to explore a better alternative, it is important to understand the current state of testing in US education and review the path that led our current situation. We are currently mired in the confusion created by the supposedly democratic practice of allowing each of the 50 states to decide its own educational standards and practices. I argue that it is absolutely necessary to create some common understanding of what student achievement across all US public school students at selected grade levels should be. Further, I argue that it is unlikely that such a welter of practices, standards, and tests as currently exists can support any real change in learning outcomes for children.

Recent economic events make clear that such a shift in student performance is imperative for everyone's future in the US: the US must earn its way back to solvency and economic power. It cannot do this without serious commitment to educating children to be thinkers and learners, not merely test takers. And right now, we do not even have the means to make progress on this front because we have no agreement about what and how to teach children at each level of their K-12 education, about what expectations for demonstrating learning should be common across all US school children at each level, and about how we should use assessment to help ensure this common standard of content and performance.

I have used the gnomic utterances of Yogi Berra to introduce the major sections of the chapter. These seemed appropriate not only because of what they say, but because the game of baseball is not entirely unlike the enterprise of education. It is absolutely dependent on team effort, it is leisurely in both its individual games and its very long season, it has very long periods of apparent inactivity (non-fans call this unbearable tedium) interrupted by moments of great drama, money for players does not guarantee outcomes and winning, and some surprisingly stolid and consistent efforts result in victory. In addition, the enterprise of baseball is preoccupied with data about individual and team performance, and with interpretations of those data to increase positive outcomes. The differences in the use and agreed upon purposes of these data between baseball and education are salient: in baseball, everyone agrees on what success looks like, and everyone acknowledges that 100% success for everyone in the enterprise is impossible. Perhaps more important, in baseball using the data for specific purposes of improvement in specific areas is quite sophisticated and disciplined. I think educators could learn a great deal from reading a book like *Moneyball: The Art of Winning an Unfair Game*, by Michael Lewis (2003), which details the ways in which consistent and persistent interpretations of data can yield surprising performance gains even when the raw material—players and salary resources—are unpromising.

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C \_ Springer Science+Business Media, LLC 2010*

**4.1 Part I: Current State**

We made too many wrong mistakes  
–Yogi Berra

In the fall of 2010, almost 8 years after the passage of No Child Left Behind—the most far-reaching testing legislation in the history of US education—where are we in terms of awareness of and more effective deployment of connections between teaching and learning as a result of increased student testing? My answer would be, “No farther down the path than we were in 2002.” Indeed, we may be substantially more distant from the goal of improving children’s learning outcomes, because across the nation educators have focused so narrowly on testing outcomes as an end in themselves.

To be sure, we now have empirical evidence of systematic differences in the ways children of poverty, color, and disability and more advantaged children perform on standardized tests, regardless of which tests are employed in which geographies. This is not a startling new fact for educators. However, its undeniability, given the disaggregated NCLB test results is, perhaps, a positive outcome, because it makes it more difficult for even the most self-serving politician to avoid the results of the virtually universal differentiation of educational results that is tied to these different subgroups’ performances. And the spotlight NCLB has focused on achievement for *all* students, and particularly those students who have persistently occupied the lowest levels of achievement, is positive and important.

However, the imposition of unrealistic goals through Adequate Yearly Progress (AYP), the culture of punishment of schools and teachers that such target shooting creates, and the persistent refusal to recognize the disparity of expectations, standards, and test rigor across different states has led to perverse results<sup>1</sup> that have weakened the links between assessment and learning.

How did we get so mired in this confusing and ultimately destructive use of assessments? It is useful to step outside the cloud of argument that shrouds any discussion of NCLB for a moment and ask a simple question: “What is the point of testing children using standardized testing tools?” It is instructive to review the history of answers to this question, as they are revealed by testing practice in public education in the United States. As Robert Linn recounts in his 1995 William Angoff Lecture at Educational Testing Service, the links between externally mandated assessments and teaching have a long history. In 1992, the Office of Technology Assessment (OTA) report, *Testing in American Schools: Asking the Right Questions* reviewed the history of testing in American School from the nineteenth century (when Horace Mann introduced written examinations) to the present. Linn quotes the OTA report summary of the view that tests are a tool that could support education reform:

---

<sup>1</sup> 1 For example, the labeling of schools as “failing” in states like Florida, which has been working systematically to implement standards-based instruction and assessment, with rigorous expectations for student achievement, while schools in states with exceptionally low standards all achieve AYP.

*Aiming at Learning: Assessment as the Critical Link*

The idea underlying the implementation of written examinations... was born in the minds of individuals already convinced that education was substandard in quality. This sequence—perception of failure followed by the collection of data designed to document failure (or success)—offers early evidence of what has become a tradition of school reform and a truism of student testing: tests are often administered not just to discover how well schools or kids are doing, but to obtain external confirmation—validation—of the hypothesis that they are not doing well at all (Linn, 1995, p. 7).

While it may be important to use test results as evidence of the failures of classroom instruction, they have been seen as not merely messengers of the need for reform, but agents of reform itself. Why is this the case? Linn gives the following reasons:

1. Tests and assessments are relatively inexpensive. Compared to changes that involve increases in instructional time, reduced class size, attracting more able people to teaching, hiring teacher aides, or programmatic changes involving substantial professional development for teachers, assessment is cheap.
2. Testing and assessment can be externally mandated. It is far easier to mandate testing and assessment requirements at the state or district level than anything that involves actual change in what happens inside the classroom.
3. Test and assessment changes can be rapidly implemented. Importantly, new test or assessment requirements can be implemented within the term of office of elected officials.
4. Results are visible. Test results can be reported to the press. Poor results in the beginning are desirable for policy makers who want to show they have had an effect. Based on past experience, policymakers can reasonably expect increases in scores in the first few years of a program (Linn, 2000) with or without real improvement in the broader achievement constructs that tests and assessments are intended to measure. The resulting overly rosy picture that is painted by short-term gains observed in most new testing programs gives the impression of improvement right on schedule for the next election. (Linn, 1995, pp. 7–8)

For more than 25 years, beginning with minimum competency testing in the 1970s and continuing with standards-based reform efforts through the 1990s and attendant test uses, education policy makers in the United States have advocated testing as a means to demonstrate how seriously they are taking raising student achievement. Certainly some of that focus was directed at identifying individual teachers and even schools that were “failing,” but much of it was directed sincerely at efforts to improve the overall status of student learning in the United States. And from 1969 forward, the federal government, through its development and administration of the NAEP, invested in the assessment of student learning across the developmental span of public schooling (sampling at grades 4, 8, and 12), critical curriculum areas (mathematics, reading, science), and—most important—time. NAEP is the only source of long-term trend data on student learning in the United States over the last four decades.

By 1995, when Robert Linn delivered the lecture quoted above, there was a new focus in states on the nexus of performance-based assessments, content standards, and performance standards, as a result of the Clinton administration’s *Goals 2000* education initiative. Lauren Resnick led the performance-based assessment charge, articulating three basic premises that ground advocacy for this kind of assessment as truly contributing to learning:

- WYTIWYG, a play on the then-new “What You See Is What You Get”

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

description of computer screen presentation: What You Test Is What You Get.

- “You do not get what you do not assess.”
- (Assuming that some kind of testing for accountability will be a permanent feature of the education enterprise) “Make tests worth teaching to.” (Resnick & Resnick, 1992, p. 59)

Concomitant with the Resnick “new standards” approach to assessment was a national focus on the creation of content standards. The National Council of Teachers of Mathematics (NCTM) led the way, and by the end of the decade of the 1990s, national and state content standards existed in almost every field. For teachers, the National Board for Professional Teaching Standards articulated in its field by field examination of accomplished teaching the knowledge standards in virtually all fields of teaching, and these teacher content knowledge standards were coordinated with content knowledge standards across the various disciplines, such as mathematics, science, and language arts.

Then there was a preliminary skirmish with the vexing issue of performance standards. If content standards—often negotiated political settlements among warring factions in a given field of study—covered the “what” of learning, performance standards were intended to wrestle to the ground the issue of “how much.” Given the mandates of NCLB, it is instructive to review some basic characteristics of performance standards. As Linn says,

There are at least four critical characteristics of performance standards. First, they are intended to be absolute rather than normative. Second, they are expected to be set at high, “world-class” levels. Third, a relatively small number of levels (i.e., advanced, proficient) are typically identified. Finally, they are expected to apply to all, or essentially all, students rather than a selected subset such as college-bound students seeking advanced placement. (Linn, 1995)

In the context of our current challenges, the issue of performance standards is a critical one. The performance standards established for NAEP reporting have been roundly criticized as unrealistically high (see, for example, Pellegrino et al., 1999; Shepard et al., 1993). They are, however, consistent across geography, cohorts of students, and time. One of the pressing issues raised by NCLB is the enormous differences among states’ performance standards, particularly the standard of “proficient” that matters for AYP. And it should be noted that setting performance standards is by no means a settled issue in the measurement community—there are multiple methods and no agreement about which works best. It is not incidental to this problem that the search for psychometric methods that would allow comparison of the results of different tests has resurfaced in the past 5 years, and with no silver bullet solutions.<sup>2</sup>

Finally, and significantly given what has happened in the first decade of this twenty-first century, the 1990s era of standards-based reform explicitly acknowledged the importance of Opportunity-To-Learn (OTL) standards. First introduced as a critical component of any accountability legislation by the National Council on Education Standards and Testing in 1992, this was by far the most controversial part of the *Goals 2000* legislation. These standards were voluntary in the legislation, a response to the controversy the topic engendered among those who believed passionately that equity concerns should be paramount, and those who vigorously

---

<sup>2</sup> See Robert Linn (2005a) for the latest account of the difficulties and confusion that attend the effort to make mathematical models bridge gaps in content, difficulty, and methodology in different assessments. Also Barton has described the challenges in the late 1990s to performance standards setting activities (Barton, 1999, pp. 19-20).

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

resisted the incursion of the federal government into local control of education. Andrew Porter summarized it this way:

To proponents, OTL standards represent the age-old problems of equity in education. In particular, advocates of OTL standards see them as an appropriate antidote to the potentially negative effects of high stakes testing on students who, through no fault of their own, attend schools which provide an ineffective education. To opponents, OTL standards evoke all their worst fears about federal intrusion into local control of the quality and nature of education. (Porter, 1994)

As we know now, *Goals 2000* morphed into No Child Left Behind. The importance of test scores to teachers and schools, and particularly, rapid improvements in test scores against an absolute standard, increased exponentially. Measurement experts, with Robert Linn in the lead, have repeatedly explained why the provisions of the legislation cannot work mathematically—there is no hope of 100% proficiency among any large country’s student population, and certainly not a country with 55 million students in its public schools<sup>3</sup>—and do not lead to any coherent and useful body of information about student learning that can inform educators’ practice and work to the betterment of student learning. And, indeed, a flood of information shows disparities between “success” as measured by NCLB rules and by many other metrics educators, parents, and policy makers have used to assess schools, such as a child’s civic and community engagement, what children seem to know and be able to do in the world outside the standardized test, and the like.<sup>4</sup>

What all of the focus on testing has done is create a culture of testing that confuses what we might call diagnosis and treatment. In fact, over the past 6 years, even the “diagnosis” nominally available from test results has become suspect: the work of aligning assessments to content standards, and adjusting curriculum and instruction to support both content and assessment of its mastery was only just begun by 2000. It is a long and arduous process to move a state’s education enterprise—curricula, teacher practices, assessments of learning, teacher preparation—to a new platform of integrated content and performance standards, particularly if the expectations

---

<sup>3</sup> Using NAEP grades 4, 8, and 12 mathematics data as a means to illustrate the stringency of the AYP requirements in NCLB, Linn examined the rates of increases that would be required to reach 100% proficiency by 2014 in selected states. Linn notes that all of the selected states recorded increases throughout the 90’s in the percentages of students scoring at the proficient level or above in all three grades. He then says, “As measured in this metric, however, the annual increases have been quite modest, averaging a little more than 1% at grades 4 and 8, and only a half of 1% at grade 12. Based on a straight-line projection of those rates of improvement, it would take 57 years for the percentage for grade 4 to reach 100. For grade 8 it would take 61 years and for grade 12 it would take 166 years. Looked at another way, the average annual rate of gain in percent proficient or above would have to increase by factors of 4, 4.3, and 11.8 at grades 4, 8, and 12, respectively, to reach 100% by 2014. Such rapid acceleration would be nothing short of miraculous.” (Linn, 2003)

<sup>4</sup> See, for example, the October 12, 2008 article in the New York Times, “More Schools Miss the Mark, Raising Pressure” (K. Hussey), October 13, 2008 article in the New York Times headlined “Under ‘No Child’ Law, Even Solid Schools Falter” (Dillon, 2008), and the ongoing controversy on school “grades” in New York City as chronicled in the Education Week blog, Eduwonkette. And, specific to Florida, the July 8, 2008 article in the The Gradebook blog at TampaBay.com, entitled “School Grades Up, AYP Down.”

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

are significantly higher for student learning than they have been in the past. The frenzy created by the national mandate for annual testing in reading and mathematics in every grade from 3–8 has created a devolution in assessment quality, and has made the systematic work of standards-based alignment and adjustment much more difficult to accomplish.

Norm-referenced shelf tests have been decorated in criterion-referenced (read standards-based) ruffles, teachers in many states have allocated extensive class time to test drills at the expense of standards-based curricula and states with legitimate and sincere aspirations to raise the standards of student learning (like Florida, South Carolina, California, and Massachusetts) have been punished by having significant numbers of schools labeled as “failures” on the basis of AYP results. The decision by many states to make “proficiency” mean something achievable in this context—and thus not at all what students really need to know and be able to do—looks like common sense in this atmosphere of finger-pointing and labeling.<sup>5</sup>

Throughout all of this intense focus on “accountability” testing and the results of such testing, two critical supports for real improvement in public education for all students in the United States have been undermined. The first is attention to the fundamental inequities in the opportunity for students to learn across the United States public schools; these have simply been ignored. Ironically, NCLB has confirmed and publicized these inequities even as it has ignored their implications for leaving children, particularly children in poverty, behind. There is currently no conversation or debate about reviving the Opportunity to Learn Standards of the 1990s, which, whatever their inadequacies, at least tried to engage the root causes of differential performance among US school children. NCLB has confirmed the shameful differences in student performance that persist along racial and socioeconomic fault lines in US public education, but there has been no serious policy debate about what to do to change that performance profile. By implication, NCLB has made diagnosis of the persistent ills of the public education system sufficient; analytical investigation and treatment of those ills is not in the purview of the legislation.

The second critical support—clear and consistent performance standards that define both “minimum” and “excellent”—has been undermined by the proliferation of competing standards, all of which are deemed acceptable, no matter how they differ state to state. No real change can occur in public education and the educational attainment of public school students in the US without sustained and thoughtful attention to *common* performance standards. Currently in the US we have no idea what a high school diploma signifies in terms of what every student who attains the diploma knows and is able to do. We do not have even minimal commonly accepted frameworks for mathematics and reading grade-by-grade attainment. To be sure, lots of rhetoric argues for every student to take and pass algebra (for example), but such rhetoric is unconnected to systematic understanding and practice by educators in mathematics teaching from indergarten on. By insisting on the myth of “100% proficiency” and by allowing each state to decide on its own definition of “proficient,” NCLB has effectively corrupted the meaning of the terms. In

---

<sup>5</sup> For a thorough examination of the problems of the exceptionally uneven playing field created by each state defining its own performance standards, see Cronin et al. (2007) report entitled *The Proficiency Illusion*.

M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA

e-mail: mapearlman@comcast.net

V.J. Shute, B.J. Becker (eds.), *Innovative Assessment for the 21st Century*,

DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010

practice, “proficient” means minimally competent in virtually every state, though what “minimal” means varies state by state. (See, e.g., Peterson and Hess, 2008.)

These two fundamental supports of an effective national public education system that aims to improve the educational attainment of all students need to be the focus of national education policy going forward. Unflinching and sustained attention to differential access to adequate teaching and learning opportunities and some common benchmarks for what constitutes at least the “good enough” floor for each stage of public education—Pre-K through grade 8, and grades 9–12 culminating in a high school diploma—are cornerstones for both policy and implementation on the ground. This is old news. Inequities in the public education system in the US and the absence of common benchmarks for student achievement across state lines have been staples in education policy debates for 25 years.

However, we face some new challenges in the US. The demographics of the US public school student population are changing very rapidly: recent estimates suggest that by 2015, 6 years from now, some 55% of all students in US public schools will be children of poverty and/or living in households in which English is not the dominant language.<sup>6</sup> We already know that these demographic characteristics pose substantial barriers to student success in our current system of education. However, as a nation we have never faced the prospect of having the *majority* of our public school students characterized by these demographic variables. The long-term implications for the pipeline of students into post-secondary education and skilled jobs and careers are sobering. If we do not direct our resources into acting on what test results tell us about the unequal playing field that constitutes US public education, then we face a future where insufficient numbers of young people entering the US economy prepared to maintain and grow the enterprises, public and private, that have assured US citizens their very high standard of living.

## 4.2 Part II: What Would Be Better

It was impossible to get a conversation going, everybody was talking too much.  
—Yogi Berra

Some revision of No Child Left Behind appears to be inevitable. Presently, however, the voices of both state and national examiners of the effects of the legislation have focused on the impossibilities of the AYP calculations, the burden of expense imposed on states by the annual testing requirements, and the mixed messages delivered by the states’ standardized tests against state standards as compared to NCLB AYP standards.<sup>7</sup> Robert Linn concludes a 2005 analysis of these mixed messages with this statement:

---

<sup>6</sup> See *Quality Counts 2009: Portrait of a Population* and “America’s Perfect Storm” for a detailed account of demographic change in the United States over the next two decades.

<sup>7</sup> In his conclusions Linn notes, “The goals established under NCLB are already unrealistic for many schools that started with low performance in 2002 and will become increasingly so, not only for those schools but for all schools as the increases in AYP targets occur” (Linn, 2005b).

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*  
*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*  
*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

If the goal for 2013–2014 remains unchanged, essentially all schools will fail to meet the unrealistic goal of 100% proficient or above, and No Child Left Behind will have turned into No School Succeeding. (Linn, 2005a, p. 14)

It is clear from all of this clamor that AYP requirements will be revisited and the calculation of the legislation's requirements somehow rationalized. Early indications of the direction intended by the new administration's Secretary of Education, Arne Duncan, suggest that the Department will attempt to link teacher evaluation systems in the states with student test results and other indicators (like high school graduation rates), and will pay increased attention to the disparities across state boundaries in the rigor of state standards and tests.

If we are to make progress toward a uniform standard of educational achievement for all students in the US public education system, thinking systemically and systematically about what really affects the education outcomes for children is essential. We have focused enormous attention on testing over the past 8 years, but the result of all of that effort has been lots of data and little real information. Results across states are not comparable, and we have come no closer to agreement on even minimum requirements for learning at each grade level, or even at benchmark grade levels (like 4th, 8th, and 12th grades). We also have not assisted teachers or those who prepare teachers in focusing their efforts on effective pedagogical content knowledge in the critical foundations of the domains. Yet without some common ground in these areas—what test results really mean, what content and skills must be mastered, how teachers can best use the time available to them to assist students to these levels—we cannot change the status quo.

For the past 8 years a recurrent and justified complaint from the states has been that the gap between the exigency of the NCLB requirements and the resources available to assist states in both better testing and reforms based on the results of the testing has made real progress impossible. That gap is no longer a fact. The US Department of Education in 2009 has at its disposal a record amount of money (some \$100 billion) to spend as a result of its share of the economic stimulus package (The American Recovery and Reinvestment Act of 2009 [Recovery Act]).<sup>8</sup> Four “assurances” that the Department asserts “... make a critical contribution to student results”<sup>9</sup> are the guiding principles for states' use of the stimulus funding:

- Making progress toward rigorous college- and career-ready standards and high quality assessments that are valid and reliable for all students, including English language learners and students with disabilities
- Establishing pre-K-to-college and career data systems that track progress and foster continuous improvement
- Making improvement in teacher effectiveness and in equitable distribution of qualified teachers for all students, particularly students who are most in need

---

<sup>8</sup> See [www.ed.gov/policy/gen/leg/recovery/index.html](http://www.ed.gov/policy/gen/leg/recovery/index.html) for an overview of the Department of Education Recovery Act stimulus funding.

<sup>9</sup> [www.ed.gov/policy/gen/leg/recovery/index.html](http://www.ed.gov/policy/gen/leg/recovery/index.html), ARRA.pdf, slide5, notes

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*  
*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*  
*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

- Providing intensive support and effective interventions for the lowest-performing Schools

All four of these principles assume some common agreements that are not currently visible about knowledge and skills for both teachers and students. Indeed, at present, no common basis exists on which a conversation could take place about performance standards for all students, and the content and skills basis for those standards at each developmental level. Each constituent group has its own agenda and its own fix on some part of the challenge. Teachers, who are the only constituent group with direct effects on student learning, are at the mercy of whatever state policy governs their geographical location. Decisions about what are called “performance standards” have been made on the basis of political expediency in most states. And while many sets of state standards articulate sound and desirable scopes of content and skill mastery, the alignments among these standards and the accountability tests used by a given state are often perfunctory and mechanical at best. In the past 8 years, many teachers have simply worked backward from what is tested (and how it is tested) on the annual accountability assessments to decisions about scope and sequence. Rather than selecting fundamental pillars of a domain of learning as the “must do” foci of instruction, test specifications—the product of multiple considerations about what is feasible and affordable to test, not what is necessarily most important to know—have driven instructional decisions. And teachers almost never are afforded the opportunity to really learn to use assessment as a support for their instruction.

I believe that using NAEP—with some additions and revisions—affords us an opportunity to address this urgent need relatively efficiently. Without some agreed upon starting point for discussion of what students need to know and be able to do in reading and mathematics, we will never be able to address either inequities in access to opportunities to learn or defined levels of sufficient learning at each developmental level. One way we might deploy already existing resources to create these common agreements is to think more creatively about the use of NAEP assessment materials and data. Suppose that we were to use assessment materials and results *already available* in some new ways, directing some of the bountiful Recovery Act funds to analysis and implementation of these tools in multiple arenas? A full-scale assault on states’ individual testing programs and the replacement of these programs with some national test is certainly doomed. However, since we have a national test in NAEP, why not think about how it might be used as a support for reform of state tests and standards?

At present, NAEP is designed to report overall trends in student achievement in the US. Data are gathered using an extremely complex matrix sampling plan, which ensures that *overall* trends in each state and each demographic category reported are accurate. What NAEP is not currently designed to do is to provide information on individual student performance, or even individual school performance. However, NAEP assessment tasks are widely acknowledged to be more sophisticated and powerful in design and evidence-gathering potential than virtually any in use in standardized accountability tests in the individual states. In addition, there is a wealth of useful information on the government’s NCES website, that includes some 2,000 NAEP tasks, state by state disaggregated data over multiple years, background variable information about students and schools that is linked to state performance on NAEP, and much more.

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

How might NAEP assessment tasks and results be used for a new purpose focused on bringing all states up to a common standard? First, some analysis of the connections among the NAEP Frameworks for mathematics and reading at 4th, 8<sup>th</sup> and 12th grade levels, and states' own standards and tests must be completed. Such an analysis could indicate the extent of the alignment between NAEP and the states' content and performance standards. Second, serious rethinking of the NAEP performance standards and benchmarks in light of their potential use as national standards and benchmarks for all states and students should be undertaken, and those performance standards revised.<sup>10</sup> Third, thoughtful analysis of NAEP tasks and their connections to the domains of instruction should be undertaken *by teachers working with NAEP developers*. Clearly, a cross-analysis of the 50 sets of state content standards in mathematics and reading would converge on the critical framework for this instruction at the three grade levels.

Fourth, a set of equivalent tests of NAEP tasks sufficient for valid and reliable individual student reporting at 4th, 8th, and 12th grade levels in mathematics and literacy (reading and writing) should be assembled and used for all US public school students to establish a baseline for achievement across all US public school students. And fifth, the implications of NAEP tasks and performance levels for teacher preparation should be explored and applied by all teacher preparation institutions so that new practitioners have a concrete sense of the progression in student knowledge and skills that is expected across the K-12 continuum.

All of these steps are possible, and certainly steps one through three could be completed by 2011, the deadline for use of the non-renewable Recovery Act funding. And the cost of national benchmark testing for every student in 4th, 8th, and 12th grades would be much less than the cost of developing new state tests for every grade. More importantly, the value of using NAEP differently is in its power to establish connections across grade levels, across standards and performance benchmarks, across critical pieces of content domains. To put it differently, using NAEP as the foundation for comparable information about student learning in this fashion makes it possible to place assessment at the center of *learning*, its rightful home from which it is currently exiled.

Numerous obstacles must be recognized and then removed for the systematic use of NAEP results to serve as the basis for comparison not just of overall state by state performance levels, but of individual student achievement. These obstacles are not immovable, and the promise of such an approach beyond formal interstate comparisons is very great. That such an existing wealth of assessment tools could serve as the foundation for an effort to integrate assessment—not just testing—with teaching needs to be at least seriously considered as we move forward to improve NCLB.

---

<sup>10</sup> For some insight into the evolution of the current “Basic/Proficient/Advanced” NAEP performance categories, and the reasons they cannot be used as currently designed for any national reporting of student-level results, see Bracey (2008) and Pellegrino (2007).

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C \_ Springer Science+Business Media, LLC 2010*

**4.3 Part III Practical Solutions**

In theory there is no difference between theory and practice. In practice there is.  
 –Yogi Berra

To demonstrate one specific positive consequence of implementing NAEP-based national benchmark testing, we need to think first about instruction, not testing. The most far-reaching and fundamental changes to the status of student achievement cannot be imposed by improving current assessment practices alone. If we do not inextricably ally assessment with instruction—as in what happens daily in each classroom—we will continue to enforce the separation of “tests” from instruction, and teachers will continue to see the tests as an external judge and jury. As Daniel Koretz pointed out in a recent conversation with an *Education Week* reporter, no one is paying much attention to the locus of instruction, the classroom. He said, “There is nobody involved in this system who has an incentive to look for good instruction anymore—all the incentives are lined up in one direction: Increase scores on the summative tests at any costs. We need to create a system in which somebody... has incentives to make sure we’re not just gaming the system.” (Cech, 2008, October 1]. Testing Expert Sees ‘Illusion of Progress’ Under NCLB. *Education Week*. Retrieved September 19, 2009 from <http://www.edweek.org/ew/articles/2008/10/01/06tests.h28.htm.?qs+Scott+Cech+October+2008>)

Exploration and investigation of explicit connections between assessment and instruction, which are critical to effective teaching and, thus, student learning have been peculiarly absent in the past 8 years. Instead, there has been an effort to make formal testing the most important influence in educational practice. The use of a single standardized test score as the driver not just of instructional practice and curriculum, but also the driver of judgments about teachers’, educational leaders’, and schools’ effectiveness has weakened considerations of what actually happens, day to day, in schools. Every day, teachers plan and preside over instruction. Effective teachers do this with constant awareness of what is happening as a result of the instruction—they observe students and their responses and learning, they create opportunities for more demonstrations and practice. This is assessment.<sup>11</sup> Formal testing is one of many assessment modalities, and the annual standardized testing of students en masse is the assessment modality most removed from instruction.

Replacing the current culture of *testing* with a culture of *assessment* could create a fundamental shift in the ways educators approach the work of teaching and learning. There is no doubt that summative testing is here to stay. However, far more significant to the real work of education, which happens *in the classroom*, is the much larger set of activities that constitute learning and socialization. Assessment is a constant companion and support to effective instruction. Putting the principles of good assessment at the center of our preparation of new teachers, as well as the ongoing professional development of practicing teachers (and principals) could radically alter not only instructional practice, but also the attitudes and beliefs about

---

<sup>11</sup> For a thoughtful discussion of the power of such formative assessment see, for example, Black and Wiliam (1998).

*Aiming at Learning: Assessment as the Critical Link*

student learning that practitioners develop and deploy.<sup>12</sup> Currently, however, our teacher preparation institutions and programs, domain-specific practitioner organizations (like National Council of Teachers of Mathematics and National Council of Teachers of English), teacher organizations, and policy organizations do not even agree about a vocabulary for discussing essential performance standards for students, and the content and skills on which those performance standards rest.

Broadening the use of NAEP, so that the domain framework and attendant assessment tasks become common ground for all teachers in all states (connected to each state's standards in explicit crosswalks), could create a basis for a national conversation and collaboration about the intersection of teaching, learning, and assessment. Analysis of the progressive knowledge and learning connections in mathematics and literacy from grades K-4, 4-8, and 8-12, as they would be assessed in the benchmark NAEP-based assessments, could lead to real engagement with some critical teaching and learning issues. These include debates about learning progressions, establishment of domain-specific content and skill frameworks, and—not least—ongoing development of practitioners' knowledge about and practice of assessment in the classroom. (Note, in this connection, the existing database of some 2,000 NAEP assessment tasks—a treasure trove from an instructional perspective.) These conversations, were this common ground created, could be led by the national teachers' organizations, the accrediting bodies for teacher preparation institutions, and organizations like the CCSSO<sup>13</sup> and the National Conference of State Legislators that provide large-scale forums for education policy debate.

Specific performance benchmarks for all children across the US can begin to organize fruitful conversations about content and curriculum, and about the kinds of pedagogical content knowledge that leads to students' success in meeting these benchmarks. These periodic checks at grades 4, 8, and 12, on progress in student learning not only can yield empirical evidence of the size of the gaps, if any, between a state's own standards and accountability tests and proficiency levels, but also differential achievement in cohorts of students in different schools and districts within a state. Empirical evidence that allows for comparison across all students in the public education system against a common set of performance standards prepares the way for addressing inequities in educational opportunities for learning as well.

The kind of focused attention to common performance standards and the content and skills they embody is most important, however, because it would create an entry point for thoughtful consideration of teachers' practice. Purposeful shifts in teachers' practices—their uses of time,

---

<sup>12</sup> For a sensible and thoughtful reflection on the primacy of teachers' judgments and assessments in bolstering student achievement, see Barton, 1999, particularly pages 31–32) A fundamental shift in the preparation of new teachers and the ongoing professional development of veteran teachers would be a focus on defining the purpose for every teaching occasion in terms of what the desired learning would actually look like—that is, what measurable and/or observable evidence of learning is the goal? Once that is established, the “what shall I do” of the lesson becomes a process of designing instruction, practice, and assessment to reach that goal.

<sup>13</sup> The CCSSO has already done extensive work in this area. See, for example, a new report at [www.ccsso.org/content/pdfs/Transforming Education - CCSSO discussion document.pdf](http://www.ccsso.org/content/pdfs/Transforming%20Education%20-%20CCSSO%20discussion%20document.pdf). In addition, the CCSSO has been at work on a common core standards project for some time; see [www.ccsso.org/federal\\_programs/13286.cfm](http://www.ccsso.org/federal_programs/13286.cfm)

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

connections among parts of domains and across domains, implementation of teaching and learning strategies that have proven effective in other places with cohorts of students similar to theirs—is the real target of interest. Without analytical understanding and discussion of what goes on *in the classroom* we can look forward to few meaningful changes in student achievement. Well-designed and focused assessment tasks—as Lauren Resnick pointed out years ago—can serve as the axis around which fundamental changes in instruction can revolve. Indeed, on a national level, almost no other catalyst is as powerful as this one could be.

#### **4.4 Part IV Creating a Roadmap with a Destination**

If you don't know where you are going, you might wind up someplace else.

—Yogi Berra

The response of schools and teachers to the mandates of NCLB is instructive, and even promising. Linking accountability test results with both funding and public scrutiny is clearly a powerful motivator. What is missing at present is assessment that is worth teaching to, that allows for meaningful comparisons across the US, and that is connected with instruction in fundamental ways. Using NAEP content frameworks, tasks, and data differently offers the possibility of tests worth teaching to, and at least as important, the deployment of already existing public resources to assist in moving practitioners to a culture of assessment.

The ultimate achievement in the creation of an assessment culture among educators, what we might term “evidence-focused teaching” is very similar in its principles to the methodology for test design and development Russell Almond writes about in Chapter 4, this book. In this “evidence-centered design” methodology, ask first what it is we want to measure—or, to put it another way, what it is we want to be able to say about someone’s knowledge, skills, or abilities—and then articulate what the credible and sufficient evidence of such a claim about the person would be. Asking first what we want to be able to say about a person’s learning, and only after we have satisfactorily articulated that, asking what we would regard as credible and sufficient evidence for saying that, is what turns most instructional planning on its head.

Thinking about assessment in this way reveals how little we can say about a student’s learning status if we depend solely on standardized test results. Not only is the test itself a very limited sample of any learner’s knowledge, it is a self-contained universe: we can talk about test scores in terms of the test and in terms of other test takers, but the world of instruction and the range of our aspirations for learning dwarfs the content of any standardized test.

Making the most of the time with learners is the province of the teacher, and it is s/he who is in the best position to gather, engender, and evaluate evidence. We have done very little to maximize the possibilities and promise of this naturally occurring phenomenon. But I believe that a systematic approach to focusing teacher preparation and professional development on a culture of assessment would yield enormous benefits for student learning—and, not incidentally, this would be evident in standardized test results.

However, this great shift cannot even be begun if we do not have some clear common understanding and agreement about what learning we are aiming to foster, and how the various stages of that learning are connected. Using not just NAEP assessment tasks, but the NAEP

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

content domain and skills frameworks (however revised) and the performance standards (also revised), allows for a much more nuanced view of student learning. Creating a national assessment at three critical points in a student's schooling acknowledges a fact of human learning that annual standardized tests do not: learning is not linear and its effects resist efforts to discipline them into organized bits. Learning takes place in ill-structured fits and starts over a time continuum much less tidy than a school year.

Common benchmarks and content domain frameworks allow the professional *selection* of instructional strategies that have a much larger and more nuanced target than answering a sufficient number of this year's annual accountability test multiple choice questions correctly. If the most important goal in the NCLB legislation became a state's demonstration of real progress against the national benchmarks at 4th, 8th and 12th grade levels for all students, what schools might accomplish, and the ways they might go about accomplishing that could radically change. Not only could teachers plan and use data *across* grade levels, they could also begin to see the implications of mathematics and reading domain frameworks through all the content they teach.

Making assessment the axis of effective teaching, and preparing new teachers to think of assessment in this fashion, is fundamental to real change in student learning. Having a central critical focus—all of this is necessary, but not necessarily sufficient—would, at the very least, make it possible to begin to talk about evidence-centered teaching. It is a very unusual teacher who plans her assessment *before* s/he plans instruction. Indeed, most formal classroom assessment is an afterthought, and most teachers are ill-prepared to think systematically about assessment design and evaluation of assessment results. If, however, the first question teachers asked of themselves when planning instruction were “What will be the observable differences in what Jonathan, Rashid, Elena, or Sung can do after I teach  $x$ ” instead of “What shall I do (usually meaning activities, assignments, content) tomorrow, this week, in this unit to teach  $x$ ?” the clarity about the focus of each session with children would be greatly enhanced. This is so because knowing what you are looking for increases the chances of finding it.

In the case of teaching and learning, framing with some specificity what the observable evidence of students' learning in a particular topic or subject area should be generates the activities that will lead to that evidence. Having a clear sense of what learning looks like, what progress for a particular learner would be, can lead the teacher directly to those activities, assignments, and exercises that would yield that kind of evidence. And knowing what the ultimate goal—what *must* students know and be able to do by a certain grade level—makes it possible to focus attention and time effectively. No one who has seen the four-inch binders of state content and skill standards given to teachers at every level would deny that such focus and selection in instruction is important.

Most teachers, particularly beginning teachers, begin in instructional design by deciding what they will “do” in a particular period of instruction. This is also the place much assessment design begins. Designing the task first, and figuring out later—if at all—what its purpose might have been or what evidence it may have yielded about the student's learning is an almost universal practice among teachers (and, truth be told, professional test makers). Changing the order of steps in instructional design, so that designing the actual activities that will take place during the

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010*

*Aiming at Learning: Assessment as the Critical Link*

lesson or sequence of lessons comes last is the single most powerful change that can be made to transform teaching. Control over what to do in class comes from analysis of why you are doing it and what you want to be able to say about its results.

What will we *do* during class? What kinds of activities should I get ready for tomorrow? How can I engage the students in this topic? What would be fun to read? How can I keep them occupied and engaged? What does the curriculum say comes next? These are the usual drivers of instructional planning for teachers.

If, however, teachers first figure out the purpose of each activity, and the learning goals each will further, and only then decide what to *do*, all of these questions become much less urduensome to answer. What activities the lesson should include, what the teacher should provide by way of instruction and resources, and—most importantly—what students should do and produce, is indicated by the teacher’s analysis of what evidence she needs to support the claims she wishes to make after the instructional period is over.

The activities done in the classroom, the homework, the assignments—all of these produce evidence about student learning if they are carefully designed. And evidence of learning is exactly what the teacher wants to observe, encourage, and gather. If evidence of increased vocabulary is needed to support the claim to be made about a particular student or group of students, then clearly some activities, assignments, and resources devoted to developing vocabulary—and observation of how each part of the planned learning is progressing—are at the heart of the teacher’s work with these children.

Assessing the status of a learner’s vocabulary at the beginning of the period of instruction is stage one in the evidence-gathering process. Designing learning opportunities that will move him/her forward toward the desired claim becomes a shaping influence on the lesson design for the instructional period. And essential to the claim about the learner’s vocabulary at the end of the instructional period is the teacher’s provision of opportunities to *see* evidence of each learner’s progress as a learner.

Imagine the effect if such an analytical planning process were oriented by common understanding of the critically important indicators of sufficient progress at selected developmental levels! Currently, most teachers experience their marching orders regarding curriculum as “everything not forbidden is compulsory.” The only real guideline right now is the state accountability test, most often a single occasion, largely multiple-choice assessment that delivers a numerical judgment of the teacher’s and school’s effectiveness. We need to do better than this.

The implications of national content and performance standards that define the absolutely necessary content and both minimally acceptable and excellent performance would support not only effective instruction in the classroom, it would also profoundly alter the way teachers are prepared for the profession. While the current confused state of teacher preparation is the subject for a different paper, imagine the galvanizing effect of known and agreed-upon benchmarks for mathematics and reading at 4th, 8th, and 12th grade levels on what beginning teachers need to know and be able to do.

*M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA*

*e-mail: mapearlman@comcast.net*

*V.J. Shute, B.J. Becker (eds.), Innovative Assessment for the 21st Century,*

*DOI 10.1007/978-1-4419-6530-1\_4, C \_ Springer Science+Business Media, LLC 2010*

## 4.6 Conclusion

The Future Ain't What It Used to Be  
–Yogi Berra

The challenges that will face the public education enterprise in the US in the next two decades are formidable. In the spring of 2009, the cohort of 18-year-olds in the US population peaks, and the number of young people in the US population steadily declines for the foreseeable future.<sup>14</sup> In addition, however, the patterns of population growth are changing, and they will fundamentally alter the tasks of teaching and learning. An increasing number of public school children will come from families whose first language is not English, many of whom face the challenge of marginal incomes or outright poverty.<sup>15</sup>

The difficulties we face now, in 2010, with high school graduation rates stagnant since the 1970's (Barton, 2005), persistent achievement gaps among subgroups in the population, and mediocre showings in international student assessments, pale in comparison to those about to envelop our education enterprise. We will not fulfill the promise of free public education for all of our children, which is the promise of possibility and prosperity for anyone who is willing to work at learning in the US, if we continue to do what we have always done.

An analog for the inevitable disappointments that will attend continuing to practice testing as if it were more than a diagnostic tool—as if it were, indeed, itself an end and not a means—is the fate of technology in US schools. In a decade in which the advances in technology have transformed the world's economy, the ways people create, access, and synthesize information, and, experts speculate, even cognitive processes themselves, instructional practice in US schools has remained virtually unchanged from 50 years ago. A teacher from the 1940s would not find herself uncomfortable with the instructional tools in use in most classrooms today, even though we have spent untold dollars on computers and technology-based instructional software. The stunning absence of the effects of advances in technology on the actual daily work of teachers illustrates the power of habit and culture, of course. But it also illustrates the absence of any sustained practice of evidence or inquiry as the basis for educating children. Technology has failed to profoundly alter educational practice because we have not asked the right questions. And those questions center on the exploration of what it is we want to aim for first, before we fall in love with the projectile we have for its own sake. Shifting from a culture of testing to a culture of assessment could profoundly change both the methods and the effects of instruction (just as technology promised to do, and has done, outside the world of “school”). It will not happen without profound shifts in emphasis and policy, however.

---

<sup>14</sup> *American's Perfect Storm*, pages 19-20

<sup>15</sup> See *America's Perfect Storm* for a detailed analysis of the interaction among changes in the economy, demographic trends, and the persistent divergence in skill distribution among groups in the US population.

M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA  
e-mail: mapearlman@comcast.net

V.J. Shute, B.J. Becker (eds.), *Innovative Assessment for the 21st Century*,  
DOI 10.1007/978-1-4419-6530-1\_4, C \_ Springer Science+Business Media, LLC 2010

*Aiming at Learning: Assessment as the Critical Link*

Aiming at learning demands that we first pay serious attention to what the critical learning accomplishments are, and what they look like as they unfold, before we decide how to accomplish or measure them. This is very hard work, and we have not yet done it. Until we do, we will not succeed in altering the outcomes for children. The implications of creating a national set of benchmarks and performance standards are also profound for those who prepare teachers to enter the profession. Not only would there be some consistency in what beginning teachers need to know and be able to do in the domains of mathematics and reading, especially for K-8 teachers, there would also be the hope of shifting the culture of teacher preparation to a more evidence-centered approach to teaching and learning. This would mean that university faculty would need to alter their practices and, in many cases, bolster their own learning. Indeed, establishing common goals for all students regardless of where they live in the US, has a better chance of putting us on the road to understanding of the learning process than any other single change we could make.

## References

- Barton, P. (1999). *Too much testing of the wrong kind; too little of the right kind in K-12 education*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Barton, P. (2005). *One third of a nation: Rising dropout rates and declining opportunities*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Black, P., & Wiliam, D. (1998). Inside the black box. *Phi Delta Kappan*, 80(2).
- Bracey, G. W. (2008). Cut scores, NAEP achievement levels and their discontents: the attempts of political bodies to bludgeon public schools with arbitrary performance standards. *School Administrator*.
- Cavanagh, S. (2007, June 13). State tests, NAEP often a mismatch. *Education Week*, 26(41).
- Cech, S. J. (2008, October 1). Testing expert sees “Illusions of Progress” under NCLB. *Education Week*, 28(6).
- Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*. Washington, DC: Thomas B. Fordham Institute.
- Dillon, S. (2008, October 13). Under ‘No AQ6 Child’ law, even solid schools falter. *New York Times*.
- <http://blogs.edweek.org/edweek/eduwonkette> (n.d.).
- <http://www.ed.gov/policy/gen/leg/recovery/index.html> . (n.d.)
- [http://www.ccsso.org/content/pdfs/Transforming Education](http://www.ccsso.org/content/pdfs/Transforming%20Education.pdf). (n.d.)
- [http://www.ccsso.org/federal\\_programs/13286.cfm](http://www.ccsso.org/federal_programs/13286.cfm)
- Hussey, K. (2008, October 12). More schools miss the mark, raising pressure. *New York Times*.
- Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America’s perfect storm: Three forces changing our nation’s future*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York: W. W. Norton.
- Linn, R. (1995). *Assessment-based reform: Challenges to educational measurement*. Princeton, NJ: Educational Testing Service.

M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA

e-mail: [mapearlman@comcast.net](mailto:mapearlman@comcast.net)

V.J. Shute, B.J. Becker (eds.), *Innovative Assessment for the 21st Century*,

DOI 10.1007/978-1-4419-6530-1\_4, C \_ Springer Science+Business Media, LLC 2010

*Aiming at Learning: Assessment as the Critical Link*

- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29, 4–16.
- Linn, R. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31, 3–16.
- Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32, 3–12.
- Linn, R. L. (2005a). *Adjusting for differences in tests*. Washington, DC: National Academy of Sciences.
- Linn, R. (2005b). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives*, 13(33).
- Pearlman, M. (2001). Kissing the frog: How teachers can transform testing. *Testing: The good, the bad, the ugly*. Washington, DC: American Federation of Teachers.
- Pellegrino, J., Jones, L., & Mitchell, K. (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press.
- Peterson, P. E., & Hess, F. M. (2008). Few states set world-class standards. *Education Next*, 8(3).
- Porter, A. (1994). *The uses and misuses of opportunity to learn standards*. Brookings Institution conference, Beyond Goals 2000: The Future of National Standards in American Education, Washington DC.
- Resnick, L., & Resnick, D. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37–75). Boston, MA: Kluwer Academic Publishers.
- Shepard, L., Glaser, R., Linn, R., & Bohrnstedt, G. (1993). *Setting performance standards for student achievement: A report of the National Academy of Education Panel of the evaluation of the NAEP Trial State Assessment: An evaluation of the 1992 achievement levels*. Stanford, CA: Stanford University, National Academy of Education.
- Solochek, J. S. (2008, July 8). School grades up, AYP down. *The Gradebook @ TampaBay.com*

M. Pearlman Pearlman Education Group LLC, Lawrenceville, NJ, USA

e-mail: mapearlman@comcast.net

V.J. Shute, B.J. Becker (eds.), *Innovative Assessment for the 21st Century*,

DOI 10.1007/978-1-4419-6530-1\_4, C\_ Springer Science+Business Media, LLC 2010