

# Chapter 3

## Assessment for Learning: US Perspectives

Jim Flaitz

### 3.1 Introduction

In the USA, educational policy-making is largely a state and local matter, rather than a federal or national area of involvement. With a few major exceptions, federal engagement in education had been limited to providing supplemental funding of compensatory education programs until recently, with the enactment of the No Child Left Behind (NCLB) Act in 2002, an early initiative in the presidency of George W. Bush. The act marked the culmination of a movement that characterized the USA as being in crisis, due in large part to an educational system that was not producing a workforce with the requisite skills for economic competitiveness. Key elements of the educational reform included a call for the establishment of rigorous standards in certain “core” subjects as a means to promote excellence in education and to make schools accountable for the academic performance of their students. With the passing of NCLB, funding for state and local education in the federal budget was consolidated and made contingent upon the states adopting the framework of NCLB.

While a range of initiatives have been introduced around the world to promote assessment as a tool for enhancing student learning, NCLB has been driving practices in the opposite direction. Although systemic changes to education at the national level for the purpose of greater accountability have been a common theme in many contexts internationally, those reforms have nonetheless created a certain amount of space for pedagogically-oriented initiatives such as assessment for learning. This space is more constrained in the USA, although educators have been attracted to complementing test-based practices with alternative, formative assessments. The difference in the available pedagogical space, suggests Popham (2006a, p. 90), stems from the fact that education reform in the USA produced “test frenzy”, while the reforms in other education systems were less frenetic. Additionally, what seems most compelling in the case of the USA is the degree of politicization of

---

J. Flaitz (✉)

Educational Foundations and Leadership Department, College of Education at the University of Louisiana at Lafayette, Lafayette, LA, USA  
e-mail: jflaitz@louisiana.edu

educational reform, in combination with strong, private sector interests representing the testing industry. Shortly after the passage of NCLB, Popham offered the wry observation at a national educational research meeting that it would perhaps be more appropriate to refer to the legislation as “No Test Publisher Left Behind”, in reference to the almost total focus placed in the legislation on the use of high-stakes tests (the near exclusive domain of private test publishers) to gauge attainment of standards by students.

This chapter explores and explains the constraints imposed upon many US schools and teachers by recent reforms such as NCLB that mitigate their opportunities to explore alternative approaches to assessment. It presents a snapshot of contemporary assessment practices, in particular those that are associated with NCLB and then investigates some practices in the spirit of assessment for learning that have managed to survive or emerge as educators at the state and local level struggle to reconcile the powerful influences of high-stakes testing with the more fundamental mandate to promote learning.

### **3.2 Economic, Political and Ideological Background to Education Reform in the USA**

In their study of education and national development, Fägerlind and Saha (1989) propose a triadic framework for analyzing reform that covers economic, political and ideological perspectives. These perspectives – which are often intertwined – can provide a contextual explanation for the ideas that underpin NCLB.

A key precursor to NCLB was a document published in 1983, by the National Commission on Excellence in Education, *A Nation at Risk*. This report raised the specter of the USA losing its economic competitiveness and falling behind other economies because the educational foundations of American society were “being eroded by a rising tide of mediocrity” (National Commission on Excellence in Education, 1983, p. 1). The danger, claimed the report, was signaled by poor performances in international comparisons of student achievement, national surveys and test scores, and other indicators, and was exacerbated by economic changes that required even higher levels of educational excellence:

Knowledge, learning, information, and skilled intelligence are the new raw materials of international commerce and are today spreading throughout the world as vigorously as miracle drugs, synthetic fertilizers, and blue jeans did earlier. If only to keep and improve on the slim competitive edge we still retain in world markets, we must dedicate ourselves to the reform of our educational system for the benefit of all – old and young alike, affluent and poor, majority and minority. Learning is the indispensable investment required to success in the “information age” we are entering. (National Commission on Excellence in Education, 1983, p. 2)

A second theme that is present in the discourse of *A Nation at Risk* concerns equity across different economic and racial groups:

We do not believe that a public commitment to excellence and educational reform must be made at the expense of a strong public commitment to the equitable treatment of our diverse

population. The twin goals of equity and high-quality schooling have profound and practical meaning for our economy and society, and we cannot permit one to yield to the other either in principle or in practice. To do so would deny young people their chance to learn and live according to their aspirations and abilities. It also would lead to a generalized accommodation to mediocrity in our society on the one hand or the creation of an undemocratic elitism on the other. (National Commission on Excellence in Education, 1983, p. 7)

Attention to diversity arose from political forces such as the civil rights movement that led to the desegregation of schools in the 1960s, and the War on Poverty that also dated back to presidency of Lyndon B. Johnson.

The solution to the perceived decline in students' academic performance and to the problems of equity, according to the authors of *A Nation at Risk*, lies in the promotion of excellence, which is to be achieved by setting high standards and focusing on core subjects such as mathematics, English, history/US government, and science, while rejecting at the same time "undemanding and superfluous high school offerings" (*ibid*, p.9). As the problem is framed as being a national issue, it follows that some form of national framework of standards would be required. The argument that *A Nation at Risk* makes reflects a social and economic efficiency orientation to educational aims. According to this orientation, the role of schools is to prepare future citizens who will be economically productive (Schiro, 2008).

The ideological linkage between economic productivity and education represents a third theme that influenced NCLB, namely, accountability. The notions that schools should be accountable and that a measurable output of the education system is student performance in standardized tests, are derived from a view of education as an economic commodity existing in a marketplace (Nelson, 2007). Calls to make schools more accountable for the public funding they received had been heard increasingly since the inauguration of federal programs such as Title I in the mid-1960s (O'Day, 2002). By calling for the establishment of standards to measure academic performance, *A Nation at Risk* facilitated the introduction of a system that makes schools accountable for the funding that they receive.

The rhetoric of *A Nation at Risk* consists of vilifying the current state of affairs in US schools and then promoting a particular vision of change for which, according to the document, there was strong public support: "Of all the tools at hand, the public's support for education is the most powerful" (National Commission on Excellence in Education, 1983, p. 8), although a cynic might retort that members of the public are highly unlikely to fail to support ideas such as educational excellence. Politicians at the state and federal level seized the opportunity to nail their colors to the mast of educational reform, with the result that NCLB enjoyed broad bipartisan support in both the House of Representatives and the US Senate.

### 3.3 No Child Left Behind

The NCLB legislation established a new direction in federal policy toward public education. However it did not mandate a national curriculum or set of standards, rather it mandated that states develop and adopt standards; it did not mandate a

specific national testing scheme, rather it mandated that states develop or adopt standardized tests in literacy and numeracy of demonstrated validity and reliability. It stipulated that the academic achievement standards should include at least three levels of achievement – advanced, proficient and basic—and that descriptions of the competencies associated with each level should be provided (US Department of Education, 2002). NCLB further required that only objective knowledge should be assessed, although states were allowed flexibility in determining the types and combinations of assessment to be used, on condition that all the standards were covered in depth and breadth, and that results could be reported in terms of the standards. A further requirement was that the assessments would have to be designed so as to be valid and accessible as far as possible for students with disabilities and those whose proficiency in English was limited. Assessment under NCLB has three main characteristics: it is high-stakes, designed to serve the purpose of accountability, and based on standards.

### ***3.3.1 High-Stakes Assessment***

The USA is a relatively recent entrant into the world of high-stakes testing for judging student learning and school effectiveness. One area where high-stakes testing has a history has been in the arena of university admissions, where the practice of requiring applicants to sit for either the Scholastic Aptitude Test/Scholastic Assessment Test (SAT) or the American College Testing (ACT) tests of “scholastic aptitude” has been in effect for over 70 years (Isaacs, 2001; Lawrence, Rigol, Van Essen, & Jackson, 2002). These tests are privately developed, administered and scored, and universities use the scores, in conjunction with other relevant applicant information (high school Grade Point Average, extracurricular activities, etc.) in making admissions decisions. In early days, the use of these high-stakes, standardized assessments for university admissions were initiated at the behest of private universities and the most selective of public universities, who were looking for an objective source of evidence of academic potential that could be efficiently applied to a growing pool of applicants from very different academic backgrounds.

Ironically, the introduction of high-stakes testing to university admissions began as an egalitarian effort (to level the playing field among applicants by providing a common measure of educational readiness). As university-going exploded after World War II (due to the return to the workforce of millions of young men), universities found themselves in the dilemma of selectively admitting applicants, and turned to the use of high-stakes tests. In the 1960s and 1970s, as the societal view of higher education shifted to that of an essential prerequisite for economic opportunity, many universities began to use the standardized test scores for placement decisions (determining which students would be required to take remedial coursework in preparation for the regular curriculum, and which students would be eligible for advanced placement, exempting them from introductory level required courses, or routing them to more accelerated versions of those courses). This placement process, which may be unfamiliar in other parts of the world, is a reflection of an

abiding view of students as being differentiable based on aptitude, running into a parallel view that access to public education (including university education) is an entitlement and a critical foundation for personal economic success. Consequently, public universities were in many instances admitting students who were less prepared for university-level studies. Because these tests used for university admissions by design focus on *aptitude* (general abilities) rather than on achievement of specific academic outcomes, the direct connection to such matters as “narrowing of the curriculum” or “teaching to the test” is less clear. Because for much of the history of university-admissions standardized testing, college-going was just one of many legitimate post-secondary paths a student might take, the proportion of students taking these tests has never been as high as is the case in other nations where such tests are mandated. Many high schools developed “college preparatory” curricula, but students largely self-selected for these curricula. Until relatively recently, conventional wisdom was that “teaching” to such tests was impossible, because the focus of the tests was on aptitudes that were a life-time in the making (in more recent times, several highly successful test-preparation companies have claimed substantial success in preparing students for these tests).

In the 1980s, individual states began to develop or adopt high school graduation examinations (which, however, did not supplant existing, high-stakes tests designed for use in university admissions) and required students to attain a minimum performance on literacy and mathematics tests in order to obtain a high school diploma (Jacob, 2001; Marchant & Paulson, 2005). These tests were *minimum competency* tests in the sense that the thresholds of performance set for them were based on “minimum” expectations for high school graduates. Prior to that time, there had been only one instance (New York state) of state or local educational agencies using high-stakes tests or examinations to make decisions about student progression, graduation, or selection to university. Instead, most states were using low-stakes testing at selected grade levels to generate comparative data on students and schools, which was putatively used for student advisement and school improvement.

An important cultural artifact that underpins much of the history of standardized testing in the USA has been a widely held belief by many in the public that differences in student school outcomes are primarily due to intelligence, which is viewed as a relatively immutable characteristic (Shepard, 2000). Consequently most standardized testing done in the schools up until the 1980s held neither students (because they could not control their own intelligence) nor schools (because they could not influence student intelligence) accountable for learning outcomes. Those views became less tenable in the face of persistent “gaps” in achievement between racial groups, and less relevant when educational quality began to be indicted as the prime cause of loss of international economic competitiveness. Those views have been largely abandoned in an era of accountability for educational outcomes. However, the past thinking has left its mark on the nature of many of the high-stakes tests still in use prevalently in the USA (multiple-choice items pitched at *aptitudes for learning* as much as at *outcomes of learning*). Since the 1990s, high-stakes testing has increasingly been used as a mechanism for introducing greater levels of accountability, at the state, district, and school levels, for student achievement.

### 3.3.2 *Assessment for Accountability*

One important difference in the way high-stakes testing in the USA has developed in comparison to other parts of the world is in the purpose of the tests. In the USA, beginning in the 1990s at the state level, and more recently at the national level with the implementation of NCLB, high-stakes testing has increasingly been used as a mechanism for introducing greater levels of accountability, at the state, district, and school levels, for student achievement. The recognition that the tests being used as a mainstay in the determination of school impact on student learning were neither designed for that purpose, nor validated for that use has been slow to come, and only recently have states moved to develop or adopt state-mandated assessments that are explicitly linked to the standards adopted by the state for student achievement.

As an element in judging school outcomes, test scores are used in two ways to influence school practices, first by making public the record of performance of schools on the tests, by way of a “report card” which compares the performance of schools at similar levels across districts within states, and second by putting into place sanctions and incentives for schools, based on those report cards. Thus schools identified as “needing improvement” may be required to provide supplemental educational support (SES) to students whose performance lags that of their peers. Another sanction applied to schools with repeated failure to achieve annual yearly progress (AYP) is the provision that parents may remove their students from that school and relocate them to a school of their choice. The ultimate sanction for consistent failure to achieve AYP is the restructuring or closing of a school.

NCLB mandated that states develop accountability plans, but left it to the states to design those plans. While school improvement and gains in student achievement are the intended goals of the legislation, there is very little in the language of the act that would spell out how schools and states will achieve those gains and improvements. With the strong focus in the legislation on test score performance, and the mandate that standardized achievement tests be the tool for accountability, it is perhaps not surprising that diverse assessment tools associated with formative functions for learning are not part of the testing landscape. An advocate of alternative assessment approaches, Richard J. Stiggins, argues:

Politicians routinely ask, How can we use assessment as the basis for doling out rewards and punishments to increase teacher and student effort? They want to know how we can intensify the intimidation associated with annual testing so as to force greater achievement. How we answer these questions will certainly affect schools. But that impact will not always be positive. . .

School administrators in federal, state, and local education agencies contribute to our increasingly damaging assessment crisis when they merely bow to politicians’ beliefs and focus unwaveringly on the question of how to make our test scores go up. . .

We are a nation obsessed with the belief that the path to school improvement is paved with better, more frequent, and more intense standardized testing. The problem is that such tests, ostensibly delivered to “leave no student behind,” are in fact causing major segments of our student population to be left behind because the tests cause many to give up in hopelessness – just the opposite effect from that which politicians intended. (Stiggins, 2002, pp. 758–579)

To date, the goal of most schools is simply to achieve AYP, as represented by the standardized achievement test results. Because this enterprise is in its infancy, relatively speaking, expressions of concern over “narrowing” of the curriculum, over-emphasis on test preparation, or failure of the high-stakes tests to take account of other, important, but more difficult to assess, learner outcomes have been largely confined to academics and professional education groups. Despite a wealth of evidence from past high-stakes testing efforts at the state level that revealed a fundamental disconnect between the *testing for accountability* approach and meaningful improvements in school outcomes (e.g., Carnoy, 2005; Darling-Hammond, 2004; Toch, 2006), and despite the emergence of similar evidence relevant to the NCLB impact (Nichols, Glass, & Berliner, 2005), the political sway of accountability combined with the relative simplicity and cost-effectiveness of standardized testing largely blunted those concerns while maintaining the ascendancy of standardized tests. Of course the paradox, as pointed out by Stiggins above and reinforced by Black and Wiliam (2005), is that the emphasis on accountability as the tool for raising standards of learning, and the use of standardized tests to measure learning outcomes, is the greatest obstacle to making gains in student achievement.

The final irony is that it is precisely the demand for accountability which has produced unprecedented pressure to improve education systems that is likely to be the biggest impediment to achieving that improvement. (Black & Wiliam, 2005, p. 260)

### 3.3.3 *Standards-Based Assessment*

Even before NCLB, with its emphasis on high-stakes testing, school improvement plans, and state accountability systems, many states had already embarked on significant reform initiatives in response to the Improving America’s Schools Act of 1994, which was earlier federal legislation aimed at promoting the adoption of world-class content standards by the states. A part of the reform was the development of performance standards (specifications of what students should know and be able to do in each content area at each grade level) and standards-based assessments to measure student attainment of the standards (e.g., Marzano & Kendall, 1996).

Direction for these reforms was taken in many states from the work of the National Research Council, which offered this description of a successful standards-based assessment system:

Research suggests that a successful system of standards-based assessment is coherent in three fundamental ways. Curriculum, instruction, and assessment all align with the standards, targeting the same goals for learning, and working together to support students’ developing understanding (**horizontal coherence**). All levels of the system (classroom, school, district, state) possess a shared vision of the goals of education, of purposes and uses of assessment, and of the criteria for competent performance (**vertical coherence**). Finally, the system needs to take into account how students’ learning develops over time. Learning progressions, descriptions of successively more sophisticated ways of thinking about an idea and laying out in words and examples what it means to move toward more expert understanding establish **developmental coherence** (National Research Council, 2005, emphasis in original).



In most states that were early, voluntary, adopters of the standards-based initiative, these assessments were intended to be low-stakes tests for both students and schools, providing information with which instructional decision-making could be guided. However, with the implementation of NCLB, achieving world-class standards, and using mandated, high-stakes examinations to hold schools accountable for achieving those standards became the context within which almost all testing took place. Although states continue to have a degree of autonomy in establishing their own content standards, the National Technical Advisory Council (NTAC) advises the Secretary of Education and the Assistant Secretary for Elementary and Secondary Education on matters relating to the approval of the design and implementation of standards by individual states. Nonetheless, significant variations in standards occur across states. A study of the standards in three states – California, Georgia, and Pennsylvania – showed that there were differences, *inter alia*, in the content of the academic standards, the difficulty level of their performance standards, and the methods for calculating AYP and AYP trajectories, and that these differences stemmed from contextual factors that pre-dated NCLB, such as the state's prior use of similar standards-based systems (Hamilton et al., 2007). The scope for variations has the advantage of allowing states to take local contextual factors into account but it complicates the national standardization goals of NCLB.

### **3.4 Assessment for Learning Under NCLB**

It is clear that, although NCLB is not mandatory in the states, its provisions for rewards and punishments, including the threat of withdrawing federal funds from non-compliant states, has brought about a systemic framework that, at least potentially, works against the principles of assessment practices that are concerned with formative functions. In a system of high-stakes testing where schools rather than students are the focus of the testing, and are the ones being held accountable for the results of the tests, the need for school-based, formative assessment as a component of the testing program has not been identified as a priority. Indeed, when the impetus behind the testing program is a suspicion that schools have somehow failed to fulfill their mandate to provide quality educational experiences to their students, and must be held accountable through the results of these tests, it should not be surprising that schools and teachers would not be seen as appropriate partners in the enterprise.

Alternative assessments, such as teacher observation of students, portfolios of student work produced during regular classroom instruction, and student performance in standardized tasks, are recognized by the NTAC for students with particular cognitive learning difficulties, provided that such assessments are aligned with state standards. Otherwise, there has been little scope within NCLB frameworks for alternative assessments. In an analysis of state assessment programs conducted in 2001 (Goertz, Duffy, & Le Floch, 2001), local assessments were found to be relatively rare among the states, and for most states that had any local assessment component, the component was most often a standardized achievement test



adopted at the district, rather than the state level. However, one direct response to NCLB has been the emergence of school-based assessment to take on that very role, primarily as a mechanism for identifying those students at greatest risk of “failing” to make AYP on the end-of-the-year standardized tests (Black & Wiliam, 2005; Olson, 2005; Popham, 2006a, 2006b). Also referred to as *benchmark assessments* or *interim assessment systems*, in many instances these “formative assessments” are simply repackaged versions of the end-of-the-year high-stakes tests intended to be administered at various times earlier in the year to detect students who may be lagging (Popham, 2006b). Increasingly these tests are being developed and marketed by third parties, and have come to represent a growth industry, as schools, desperate to achieve their AYP goals, search for whatever means may be available to them to raise student test scores.

However, as Popham (2006a, 2006b) points out, in many instances the diagnostic utility of these tests is very limited, and their use in genuinely formative fashion is rare. Quoted in an article by Olson (2005), Robert Slavin observed, “If you’re looking, as you should be, at the full range of development that you want kids to engage in, you’re going to have to look at their work products, their compositions, their math problem-solving, their science and social-studies performance.” While the practice of interim assessments appears to be growing among the schools (particularly schools struggling to achieve AYP), and does in a general way represent school-based assessment, it fails in several important respects to accomplish many of the aims associated with the concept in other parts of the world (broadening the base of curriculum being assessed, promoting assessment for learning, contextualizing assessment of performance skills in authentic settings, embedding assessments within the learning experience). What it does seem to represent is at best, an alignment of assessments occurring in the school setting with those being administered as the accountability measures, designed to support student achievement of learning outcomes represented on the high-stakes tests.

While most states opted for the use of traditional testing instruments (standardized achievement tests, multiple choice item formats), some states chose to employ less-traditional assessment systems, including the use of portfolios and performance assessments embedded within the learning activities in the schools. Cromey and Hanson (2000) reported on schools in Michigan, one state that had begun the process of reforming its schools several years earlier. In their study, two groups of schools were selected, one group that had well-developed school-based student assessment systems and a second group, matched on important school and student characteristics, that had less well-developed assessment systems. The purpose of the study was to identify those features of the schools with well-developed systems that distinguished them from the schools with less well-developed systems. Several significant differences were noted. The schools with well-developed systems:

- aligned their local curriculum, standards, and assessments to the state content standards
- analyzed assessment results to monitor student progress.
- used state assessment results to check the validity of local assessment systems

- used assessment results to evaluate the efficacy of local curriculum and instructional practices
- limited the number of student assessments used to those that are purposeful and can be aligned to local curriculum and state standards
- allocated time for teachers to collaborate, reflect and make data-based decisions- individually or in teams- based upon student assessment data and their instructional implications.

In some states high-stakes tests for high-school graduation have been supplanted by end-of-course (EOC) examinations. The best-known EOC examinations are the Advanced Placement, New York Regents, and International Baccalaureate examinations. Fifteen states include, or will soon include, EOC examinations as part of their high school assessment systems. Unlike traditional high school graduation examinations that might be administered at any grade level from grade 10 to grade 12, and might focus on learning outcomes associated with grades 8 through 12, and unlike standardized achievement tests that might focus at a specific grade level, but would typically assess across a range of subjects with relatively few items per subject, the EOC examination is a purpose-designed examination that, as its name suggests, is administered at the end of the course, making it grade and subject specific. However, like the exit examinations, they are administered by a third party rather than the teacher and, as such, allow comparability among schools and courses. Advocates of such tests argue that they offer a better basis for judging student achievement in the various subjects included in the high school curriculum, and unlike the other test formats, can provide a more valid source of evidence of student achievement in the particular subjects taken. While EOC examinations represent an approach taken in some states to address some of the problems found to be associated with traditional high-stakes testing, they nevertheless fall short as “school-based assessments” in most important respects. Although they can serve to “broaden” the curriculum by testing in many subjects, and can represent tasks that involve more complex learner outcomes than are commonly found on traditional standardized achievement tests, they are not carried out by teachers, in schools, and the tasks which are set for students are not embedded in their learning, and consequently cannot serve as the basis for supporting learning and providing constructive feedback on that learning.

One state that has developed a form of school-based student assessment modeled on those principles is Vermont. Quoting from the Vermont “Core Principles” document,

The Vermont School Quality Standards call for a balance of both classroom based and school based assessment. At the classroom level, formative assessment reflects individual student “learning in progress”. Beyond the classroom, the school based system needs to generate feedback that enables teachers and other members of the educational community to determine consistency in meeting shared expectations for student learning across all classes and grade levels. A comprehensive assessment system encompasses both classroom and school-based assessments. (Vermont Department of Education & Standards & Assessment, 2006, p. 1)

Another important feature of the Vermont School Quality Standards is the role of collaboration. To again quote from the core principles document,

Most importantly, teachers, administrators, and other members of the school community need to engage collaboratively in the decision making process, with ongoing dialogue about the relationship between learning goals and assessment. Finding the time and opportunity for collaboration is a significant challenge for a school. When such dialogue becomes part of the school's culture, significant rewards come in the form of continuity, professional development and improved student learning. (Vermont Department of Education & Standards & Assessment, 2006, p. 1)

Perhaps the most visible example of a state committed to local, school-based assessment is Nebraska (Nebraska Department of Education, 1999), with its STARS (School-based, Teacher-led, Assessment & Reporting System). Nebraska was the only state to successfully resist the NCLB mandate to base accountability decisions on nationally recognized standardized tests of achievement, largely due to its aggressive moves to demonstrate that its existing system of school-based assessments were capable of producing valid and reliable evidence of student achievement (Roschewski, Isernhagen, & Dappen, 2006). Nebraska's 517 school districts design their own assessment systems which include a portfolio of teachers' classroom assessments, district tests that measure how well children are meeting locally developed learning standards, a state writing test and at least one nationally standardized test. The last component was not originally part of the STARS program, but was instead part of the compromise Nebraska reached with the US Department of Education which allowed it to retain its assessment and accountability system. Nebraska teachers at the district level worked to align their curriculum and assessment practices to the state content standards. Standards-based classrooms are achieved as each teacher clearly articulates the learning targets, aligns instruction to the learning targets within each of the content standards, and assesses whether or not students are meeting the targets outlined by the content standards.

To ensure quality in the locally developed and administered assessments, each year a District Assessment portfolio, which includes a sample of actual assessments used in the classrooms at each grade level, is assembled and submitted to the Nebraska Department of Education. Department of Education personnel, working with consultants from the Buros Center for Testing, examine the assessment materials and evaluate them against 6 quality assessment criteria. In a recent evaluation of the STARS assessments, Brookhart (1999), noted that generally the alignment of teacher developed assessments to state content standards was good, although the reliability of teachers' judgments of student outcomes using those assessments was uneven, with teachers able to reach consensus in mathematics more consistently than in reading (the two areas assessed).

### 3.5 Recent Developments

For at least the past 30 years, changes in educational policy in the USA, particularly at the national level, have been driven by ideological views of the proper role of the national government, as well as the proper role of public education. Many of the mandates associated with NCLB reflected a conservative ideology that tended to view public education and its employees as fundamentally flawed

and responsible for many of the failings in American society. With a significant change in national leadership (in both legislative houses as well as the executive) in the most recent rounds of national elections, some of that ideology is beginning to shift. While support for the major tenets of NCLB seems to remain relatively solid, there is evidence of at least some rethinking of the rigidity of some aspects of the assessment component of the legislation. One such shift has been the move by the Department of Education to allow more states to propose alternative models for achieving the aims of the legislation. Although the reliance on high-stakes testing aligned to state educational standards remains a constant, such alternatives as end-of-course examinations are being allowed.

Another major development has been the proposal of a “value-added” model for judging student achievement and school/teacher impact on student learning (Shurtleff & Lored, 2008). The value-added model essentially uses a sophisticated statistical modeling approach to predict the test-score performance of each child, allowing for the impact of individual differences as well as past test performance. By “controlling” for those factors, the argument is that it becomes possible to detect the impact of instruction on the performance of the students on the test. Of course, all this arises from the underlying purpose of finding a way to hold individual teachers accountable for their impact on the learners in their classes, and so, unfortunately, it has proved controversial among educators.

Finally, in 2009, the US Department of Education introduced a new initiative referred to as “Race to the Top”, which provided a multi-billion dollar competitive grant to the states to promote educational innovations and reform (US Department of Education, 2009). While the purposes of the grant are to encourage educational excellence and the improvement of underachieving schools, one controversial provision has been the requirement that states tie teacher pay and retention to the performance of their students on the state’s high-stakes tests. “Race to the Top” and “value-added assessment” actually work hand-in-hand, as the one represents the means whereby the other is proposed to be achieved.

### 3.6 Conclusions

Even before NCLB, with its emphasis on high-stakes testing, school improvement plans, and state accountability systems, many states had already embarked on significant reform initiatives aimed at promoting the adoption of world-class content standards by the states. Even in those states like Vermont and Nebraska, where significant commitment to an assessment system that performs diverse functions has been in evidence, the use of the assessment results is problematic (because they are being used to judge the quality of the school and teachers, in addition to appraising the competence of the students) even if the rationale (assessment that supports learning) and the assessment approach (portfolios, performance assessments) may represent shared qualities. Unfortunately, these examples are the exception, with more evidence of states and school districts rushing to adopt “formative

assessment” systems that represent little more than “early warning systems” for student difficulties in achieving AYP (Sharkey & Murnane, 2006).

Although most educational experts and assessment experts endorse the sorts of approaches taken in Nebraska and Vermont (e.g., Darling-Hammond, 2004; McMunn, McCloskey, & Butler, 2004; Shepard, 2000; Sirotnik, 2002; Wilson & Sloan, 2000), the reality is that as pressure grows on schools to make their annual yearly progress on the standardized achievement tests, the “oxygen” needed to simultaneously support the more educationally sound school-based assessment systems is likely to disappear (e.g., Mitchell, 1997).

What is perhaps most interesting to note is the comparative enthusiasm and support shown by professional education organizations and teachers in the USA for school-based, formative assessment systems. Perhaps this is in part because many US schools and teachers have had more than a decade of experience with standards-based assessment, but very limited experience with national examinations (and the attendant pressures to shape the curriculum to those examinations). More likely, the attitudes of US teachers are being shaped by the accountability focus of the high-stakes tests, which places them in the unenviable position of being held responsible for the achievement of high levels of proficiency by all learners, irrespective of contextual factors, levels of resources, or any other considerations.

It has become something of a truism in the assessment community that “assessment drives curriculum”, and the higher the stakes associated with the assessment, the more strongly the assessment will determine the priorities of schools, teachers, and students in matters of curriculum choices, instructional practices, classroom assessment approaches, and learning strategies. So long as the stakes associated with performance on public examinations are as high as they are, it will be difficult to create a climate conducive to meaningful school-based assessment. Classroom teachers are in some cases retreating from assessment for learning practices (often at the direction of school leaders) out of a fear that they will not be preparing students for the types of assessment and learning outcomes reflected on the high-stakes tests. School-based assessment, as it is conceived in some settings elsewhere, as a means of enhancing the validity of the traditional public examinations while supporting assessment for learning approaches in the classroom, is not part of the high-stakes assessment equation in the vast majority of US states.

In those few states, school districts, and individual schools where a commitment to an assessment model in which teachers collect and use evidence of student learning to support their learning, where the assessments are “authentic” and extended, and embedded in meaningful learning activities, where students actively participate in the learning, and the assessment of that learning, it seems inevitable that those schools will find themselves under greater pressure each year to focus more explicitly on achieving those NCLB mandated targets, reflected in test scores, at the expense of focusing on the learning. Because it has been demonstrated that schools that focus their efforts on increasing test scores on a specific test do typically see test score rise *on that test*, without producing a commensurate increase in the underlying learning, the practices that lead to that increase (drilling on the test/practicing test-taking, modeling classroom assessments on the high-stakes tests, de-emphasizing or

eliminating subjects in the curriculum not represented on the test, focusing instruction on the types of learning outcomes, usually lower-level, represented on the tests) will be difficult to resist, particularly for schools that serve student groups that traditionally under-perform on standardized achievement tests.

One implication of the US experience has to do with the use of high-stakes tests for holding schools and teachers directly accountable for student test performance. Tests that began as “monitoring” mechanisms eventually evolved into “accountability” tools. In the USA, much of the impetus behind accountability testing has been political, and those winds have begun to shift with a change in national administrations. It is unlikely though that the role of standardized achievement tests will diminish significantly in the foreseeable future, nor that the role of school-based assessment, as a formal part of accountability will necessarily rise. More promising is the prospect for assessment for learning practices at the classroom level to grow as schools become disillusioned with the “quick-fix” strategies that can only produce short-term and superficial results, especially for those student subgroups that are traditionally least successful in standardized testing situations.

## References

- Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*, 16(2), 249–261.
- Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 6–13.
- Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice*, 24(4), 19–31.
- Cromey, A., & Hanson, M. (2000). *An exploratory analysis of school-based student assessment systems*. North Central Regional Educational Laboratory (ERIC document).
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047–1085.
- Fägerlind, I., & Saha, L. J. (1989). *Education and national development: A comparative perspective* (2nd ed.). Oxford: Pergamon Press.
- Goertz, M. E., Duffy, M. C., & Le Floch, K. C. (2001). *Assessment and accountability systems in the fifty states*. Consortium for Policy Research in Education, Number RR-046. Philadelphia: University of Pennsylvania.
- Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., & Russell, J. L., et al. (2007). *Standards-based accountability under No child left behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: The Rand Corporation.
- Isaacs, T. (2001). Entry to university in the United States: The role of SAT and advanced placement in a competitive sector. *Assessment in Education*, 8(3), 391–406.
- Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99–121.
- Lawrence, I., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2002). A historical perspective on the SAT: 1926–2001. New York: College Entrance Examination Board. [http://professionals.collegeboard.com/profdownload/pdf/rr20027\\_11439.pdf](http://professionals.collegeboard.com/profdownload/pdf/rr20027_11439.pdf). Accessed 7 May 2010.
- Marchant, G. J., & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Educational Policy Analysis Archives*, 13(6), 1–15.
- Marzano, R. J., & Kendall, J. S. (1996). *A comprehensive guide to designing standards-based districts, schools, and classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

- McMunn, N., McCloskey, W., & Butler, S. (2004). Building teacher capacity in classroom assessment to improve student learning. *International Journal of Educational Policy, Research, & Practice*, 4(4), 25–48.
- Mitchell, K. (1997). What happens when school reform and accountability testing meet? *Theory into Practice*, 36(4), 262–268.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the Nation and the Secretary of Education, United States Department of Education*. Washington, DC: National Commission on Excellence in Education.
- National Research Council. (2005). *Systems for state science assessment*. Washington, DC: National Academies Press.
- Nebraska Department of Education. (1999). School-based Teacher-led Assessment & Reporting System: A planning guide for Nebraska schools. <http://www.nde.state.ne.us/>. Accessed 12 November 2006.
- Nelson, C. (2007). Accountability: The commodification of the examined life. *Change: The Magazine of Higher Learning*, 36(6), 22–27.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2005). High stakes testing and student achievement: Problems for the No Child Left Behind Act. Educational Policy Research Unit. <http://www.asu.edu/educ/eps/EPRU/documents/EPSP-0509-105-EPRU.pdf>. Accessed 20 November 2006.
- Olson, L. (2005). Benchmark assessments offer regular achievement. *Education Week*, 25(13), 13–14.
- O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72(3), 293–329.
- Popham, W. J. (2006a). Diagnostic assessments: A measurement mirage? *Educational Leadership*, 64(2), 90–91.
- Popham, W. J. (2006b). Phony formative assessments: Buyer beware!. *Educational Leadership*, 64(3), 86–87.
- Roschewski, P., Isernhagen, J., & Dappen, L. (2006). Nebraska STARS: Achieving results. *Phi Delta Kappan*, 87(6), 433–437.
- Schiro, M. S. (2008). *Curriculum theory: Conflicting visions and enduring concerns*. Los Angeles: Sage Publications.
- Sharkey, N. S., & Murnane, R. J. (2006). Tough choices in designing a formative assessment system. *American Journal of Education*, 112(4), 572–588.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shurtleff, D. S., & Lored, J. (2008). Beyond No Child Left Behind: Value-added assessment of student progress. National Center for Policy Analysis. <http://www.policyarchive.org/handle/10207/bitstreams/11781.pdf>. Accessed May 7, 2010.
- Sirotnik, K. (2002). Promoting responsible accountability in schools and education. *Phi Delta Kappan*, 83(9), pp. 662–674.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758–765.
- Toch, T. (2006). Turmoil in the testing industry. *Educational Leadership*, 64(3), 53–57.
- US Department of Education. (2002). No Child Left Behind Act. <http://ed.gov/policy/elsec/leg/esea02/107-110.pdf>. Accessed 19 July 2010.
- US Department of Education. (2009). Race to the Top Program: Executive summary. US Department of Education. <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>. Accessed 5 May 2010.
- Vermont Department of Education, Standards & Assessment. (2006). Core principles of high quality local assessment systems. [http://education.vermont.gov/new/pdfdoc/pgm\\_curriculum/local\\_assessment/core\\_principles\\_06.pdf](http://education.vermont.gov/new/pdfdoc/pgm_curriculum/local_assessment/core_principles_06.pdf). Accessed 16 November 2006.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181–208.