A BRIEF TOUR OF ASSESSMENT-RELEVANT FEDERAL STATUTES

Because decisions about if or how our federal government ought to influence America’s educational assessments should surely take into consideration what has gone on legislatively in years past, it will be instructive to review, albeit briefly, the most salient federal statutes bearing on U.S. educational assessment. In turn, then, a once-over-lightly treatment follows of the National Defense Education Act, the Elementary and Secondary Education Act, the establishment of the National Assessment of Educational Progress, and the Education for All Handicapped Children Act.

Whereas it may be true that “what’s past is prologue,” past events are far from perfect predictors. A particular federal statute that could have a positive impact on the nation’s schools if administered deftly by insightful federal personnel might prove disastrous when implemented ineptly by less talented federal officials. Federal laws are enacted at a particular time in a particular social context, and administered by particular people. As we look back at previous federal legislation related to educational assessment, we may encounter some statutes that appear to have been less successful than their supporters had anticipated. This does not signify that all similar legislation is destined to stumble. We should surely try to learn from history, but not all legislative history is laden with exportable lessons. The

National Defense Education Act (NDEA)

A threatening Cold War served as backdrop for the enactment of the 1958 National Defense Education Act. It is generally conceded that the Soviet Union’s launching of Sputnik finally confirmed an increasingly widespread belief that America was technologically falling behind the Soviets. To cope with this perception, our public schools were viewed as both the problem and the remedy (Anderson, 2005). Educational assessment, however, played only a small role in this legislation. Title V of NDEA provided grant monies to states for “guidance, counseling, testing [emphasis added] and identification of able students” (Spring, 2008 p. 402). In addition, modest funding was provided for states to develop data-gathering and reporting systems (Spring, 2008). Otherwise, NDEA appears to have had little impact on U.S. educational assessment.

The Elementary and Secondary Education Act (ESEA)

In marked contrast, the Elementary and Secondary Education Act has dramatically altered the federal government’s role in U.S. public education, especially with regard to the assessment of student achievement (Anderson, 2007). ESEA has been a significant part of America’s educational landscape since its enactment in 1965 (Anderson, 2005; Jennings, 2001), and has been reauthorized eight times (Anderson, 2005). ESEA has provided billions of dollars for the nation’s public schools.

---

Congress first authorized ESEA as a part of President Lyndon Johnson’s War on Poverty (Jennings, 2001). The law’s purpose was to “provide financial resources to schools to enhance the learning experiences of underprivileged children” (Thomas & Brady, 2005, p. 51). At the beginning, and until its 1994 reauthorization as the Improving America’s Schools Act (IASA), educational assessment in ESEA was largely limited to a form of fiscal monitoring in order to ensure that federal dollars were effectively serving the intended populations (Anderson, 2007; Jennings, 2001; Riddle, 2001).

The original 1965 law contained several provisions, or “titles,” but Title I was by far the most recognized and influential. Title I was the provision supplying categorical funding for low socioeconomic status (SES) school-aged children. The law did make an attempt at evaluation (Merkel-Keller, 1986), noting that state education agencies (SEAs) and local education agencies (LEAs) could only receive Title I funds if there were “procedures for evaluating the effectiveness of the programs . . . in meeting the special education needs of educationally deprived children” and that the “evaluations will include effective measurements of educational achievement in basic skills” (Title I, Section 172 cited in Merkel-Keller, 1986, p. 8).

Section 172 of ESEA also included a reporting requirement whereby SEAs were to “make periodic reports . . . evaluating the effectiveness of the program” (Title I, Section 172, cited in Merkel-Keller, 1986, p. 8) to the Secretary of Education assuring that Title I students were making progress. Difficulties arose in that LEAs and SEAs had no consistent way to assess students, collect data, or report the findings. Nor were there adverse consequences for failing to do so, thus rendering the provision ineffectual. However, this evaluation provision of ESEA could be regarded as a significant step toward today’s federal assessment, evaluation, and accountability requirements.

The 1974 reauthorization of ESEA saw assessment gain ground. Concerns over the accounting and monitoring of the use of funds receded, and Congress turned its attention to the impact of ESEA on student achievement (Borman, 2000; Borman & D’Agostino, 2001; Borman & D’Agostino, 1996; Merkel-Keller, 1986). The 1974 amendments included funding provisions for the Secretary of Education to develop an independent and systematic evaluation and reporting system to determine the educational effectiveness of ESEA with “objective criteria . . . utilized in the evaluation of all programs . . . producing data which are comparable on a statewide and nationwide basis” (Education Amendments of 1974, 88 Stat. 484, 500 as cited in Borman & D’Agostino, 2001, p. 41). Between 1974 and 1978, the Title I Evaluation and Reporting System (TIERS) was developed. TIERS implementation began in 1979 after the 1978 reauthorization of ESEA.

TIERS consisted of a three-level process for evaluating and reporting (Borman, 2000; Borman & D’Agostino, 2001; Merkel-Keller, 1986; Rutherford & Hoffman, 1981). At the first level, TIER I, the local education agency collected student achievement data and reported this information to the state education agency. In TIER 2, the SEA aggregated the LEA data and reported these results to the commissioner of education. For TIER 3, the U.S. Office of Education aggregated the nationwide data and reported on the effectiveness of Title I programs in meeting the needs of the identified disadvantaged population.

The implementation of TIERS generated major methodological issues. For example, there were concerns about the accuracy of the cumulative national data. Because SEAs and LEAs retained control over the selection of achievement tests, there were a number of different tests employed throughout the states. To illustrate, states used such assessments as the Stanford Achievement Tests, the Comprehensive Tests of Basic Skills, and the Iowa Tests of Basic Skills. Multiple tests, of course, led to questions regarding the legitimacy of between-state comparisons. Concerns over the administration of the tests also began to
surface. Were guidelines for administration the same across all schools in an LEA as well as throughout a state? There were also questions regarding the timing of test administrations, because SEAs and LEAs often tested students at different times of the year. Although federal officials attempted to address these methodological issues in various ways, many of those technical solutions proved to be less than satisfying.

In the 1988 Hawkins-Stafford Act reauthorization of Title I, accountability requirements were further strengthened. The act required states to “define the levels of academic achievement disadvantaged students should [emphasis added] attain in schools receiving Title I funds” (Jennings, 2001; McDonnell, 2005). The 1988 version of ESEA required SEAs and LEAs not only to develop plans for increasing student performance, but also to test and report student achievement of the Title I population annually (Riddle, 1989; Thomas & Brady, 2005). While many states had by this time implemented annual assessments of their students, this was the first time that yearly administrations were required (Thomas & Brady, 2005). Moreover, there was a new evaluative provision for those schools that elected to use their Title I funds on a schoolwide basis. This provision required that after a three-year implementation, such schools must demonstrate that the achievement of their disadvantaged children was higher than either the average of children participating in the LEA as a whole, or [was higher than] the average for disadvantaged children in that school over the three years preceding the schoolwide plan’s implementation (Riddle, 1989, p. 11).

IASA, in a legislative change of considerable import, shifted its evaluative focus toward a more clear delineation of educational expectations and a more comprehensive effort to assess students’ attainment of those expectations (The Commission on No Child Left Behind, 2007; DeBray, 2005; Hanushek & Raymond, 2005; Jennings, 2001). To qualify for IASA Title I funds, “states were required to adopt and implement curriculum content standards, pupil performance standards, and assessments linked to these” (Riddle, 2001, p. 2). In addition to the development of standards and assessments of students, IASA required that all students be assessed, especially second language learners and students with disabilities. IASA, a particularly significant reauthorization of ESEA, also required that student achievement data be disaggregated for statute-identified subpopulations (Fast & Erpenbach, 2004).

A major modification in IASA was its effort to move states away from reliance on norm-referenced measurement toward the use of a criterion-referenced assessment approach. Riddle (2008) notes this significance, stating these “two types of tests vary primarily [emphasis in the original] regarding how test results are analyzed, but also typically differ to some degree with respect to such characteristics as the range of questions included” (p. 3). Therefore, because IASA required that assessments be tied to standards and to determining the achievement of those standards—rather than comparing students with one another—many state officials found that they needed to move toward a starkly different assessment strategy for their state’s accountability tests. As states attempted to accomplish this change, Redfield & Sheinker (2004) note that “approaches range from adding items to available norm-referenced tests (NRTs) to align them with state academic content standards and grade-level expectations (GLEs) to creating new standards-based tests (p. 8).”

Passage of the No Child Left Behind Act, the 2002 reauthorization of ESEA, dramatically magnified the assessment requirements of IASA by adding significant stipulations regarding assessment (DeBray, 2005; Riddle, 2008; Shaul & Ganson, 2005) and by installing a series of potent sanctions for schools and districts whose students failed to sufficiently improve their scores on a state’s annual accountability tests.
Rather than being obliged to administer assessments in a grade range (e.g., grades 3-5) as had been called for in IASA, states were required by NCLB to develop and/or adopt assessments for each of grades 3-8 and once in grades 9-12 in the areas of reading/language arts and mathematics. NCLB also required that science be added for grade ranges 3-5, 6-9, and once in high school (Erpenbach, Forte-Fast, & Potts, 2003; Mills, 2008; Riddle, 2008). In a sense, NCLB took many of the requirements of IASA and put sharper teeth into them. The disaggregation of subgroup test scores in NCLB, for example, could lead to far more serious evaluative consequences than had been the case for IASA’s subgroup disaggregation. Many observers regard NCLB’s more stringent demands for subgroup disaggregation to be a particularly significant contribution of this most recent reauthorization of ESEA.

In what may turn out to be one of its more significant changes, NCLB required states and those local districts selected in the sampling plan to participate in the National Assessment of Educational Progress (NAEP) in the 4th and 8th grade in mathematics and reading (Redfield & Sheinker, 2004; Shaul and Ganson, 2005; Riddle, 2008). This was a major change, because until NCLB, participation in NAEP had been voluntary (Riddle, 2008; Shaul & Ganson, 2005). With the availability of state-by-state NAEP scores, it was apparent that NCLB-required performances of students from different states would most certainly be compared. NAEP would, in the eyes of many policymakers, become the de facto yardstick by which the success of a given state’s public schools would be determined.

The National Assessment of Educational Progress (NAEP)

Authorized and funded by Congress when initially implemented in 1969, the National Assessment of Educational Progress (often referred to as “the Nation’s Report Card”) provides information to lawmakers, policymakers, and the general public about U.S. students’ academic achievement (Bowers, 1991; Hombo, 2003; Riddle, 1998). Since its inception, NAEP has provided a “snapshot” of American students’ achievement. In its early years, NAEP was a particularly innovative assessment program, annually assessing diverse content areas with a significant proportion of its tests consisting of performance assessments. Because of NCLB requirements, since 2003 NAEP has provided results in reading and mathematics at grades 4 and 8 every two years for states and the nation. Results are not available by school or by student, because (with the exception of several large school districts) the smallest NAEP reporting units are states.

NAEP is not one test, but rather a set of three tests that each serve a specific purpose. The NAEP trend test is given in reading and mathematics every four years to student samples at ages 9, 13, and 17 (http://nces.ed.gov/nationsreportcard). The trend test has provided longterm, comparable information about nationwide student achievement since its inception in 1969. Results are reported for the nation as a whole and for subpopulations such as gender and ethnicity. The main test, given every two years in grades 4, 8, and 12, provides measurements of student achievement in the areas of reading and mathematics. As with the trend test, results are reported for the nation as a whole and for subgroups (http://nces.ed.gov/nationsreportcard). State tests, the third type of NAEP test, were added in 1990 based on a perceived need for comparable data across states (Hombo, 2003). These assessments are administered in the opposite years from the main tests in the areas of reading, mathematics, and writing.

NAEP appears to have had an impact on the nature of America’s assessments as a whole, and is thought by some to have increased the public’s confidence in testing (Bowers, 1991). For example, NAEP results have been used to evaluate the success of programs such as NCLB (Fuller, Wright, Gesicki, & Kang, 2007; NCES, 2006; Center on Education Policy, 2008) and “to study the effect of accountability on student
performance (Herman, 2007, p. 11).” NAEP’s assessment procedures have also served as models for numerous state testing programs (Campbell, Hombo, & Mazzeo, 1999).

Not all observers of NAEP are fans. For example, in a review of a 2004 anthology about NAEP (The Nation’s Report Card: Evolution and Perspectives), Stake registers strong doubts about the ultimate benefits of this oft-evaluated program:

With strong support from the measurements community, the main characters in this book about NAEP created NAEP in their own image. They wanted it to be the best that they could be. To be pure assessment, they disdained curriculum experts and philosophers. But they failed to demand validation of the assessment’s core policy. At first, the core policy was tracing performance over time, but gradually the core policy became test-based accountability. (Stake, 2007, p. 18)

At the very least, the methodological measurement refinements emerging from NAEP have surely had a considerable influence on the way many of America’s educational assessments are currently created. Given the increasing importance attributed to a state’s NAEP scores by educational policymakers, we may well see this assessment program, almost 40 years old now, attain a position of ever-increasing prominence in the way states choose to assess their students.

To illustrate, if NAEP results become regarded as the most credible way of evaluating a state’s schools, then it is not likely that at least some state officials will urge their state’s assessment personnel to “measure what NAEP measures?” Even though, as long as NAEP scores are not reported at the school level, NAEP will remain a relatively low-stakes assessment, its curricular impact at the state level could thus be considerable.

The Education for All Handicapped Children Act (PL 94-142) and the Individuals with Disabilities Act (IDEA)

The Education for All Handicapped Children Act (PL 94-142) of 1975 and its reauthorized version, the Individuals with Disabilities Act of 1990 (IDEA) has had a substantial impact on the way many thousands of American children have been educated (Itkonen, 2007; Koegh, 2007; Thurlow, Lazarus, Thompson & Morse, 2005). Until the enactment of PL 94-142, children with disabilities had limited or no access to public education (Lehr & Thurlow, 2003; Zettel & Ballard, 1979). PL 94-142, however, mandated that children with disabilities be provided a “free and appropriate public education” (FAPE) and in the least restrictive environment, which usually meant the regular classroom (McLaughlin & Nagle, 2004). PL 94-142, as had no previous federal legislation, ensured equal access to public education for all children with disabilities (Gaddy, McNulty, & Waters, 2002; McLaughlin & Nagel, 2004).

In the early years of the law, its assessment and evaluation provisions focused on compliance and monitoring (Anderson, 2005; DeStefano, 1992; Gaddy, McNulty, & Waters, 2002; Lehr & Thurlow, 2003; McLaughlin & Nagle, 2004). However, when PL 94-142 was reauthorized as IDEA in 1997 and again in 2004, its measurement focus changed from assessing how the program was being implemented to instead determining the academic achievement of the populations served (Lehr & Thurlow, 2003; McLaughlin & Nagle, 2004). Put simply, the assessment emphasis of this legislation shifted from inputs to outputs. IDEA also mandated that state assessments for children with disabilities be tied to the same state standards that had been adopted for other children in that state.

---

With the shift to assessing academic achievement, IDEA mandated that students with disabilities be included in the same general assessment as their nondisabled peers. No longer could disabled students be excluded from the general testing programs on the basis of their disabilities (McLaughlin & Nagle, 2004; Thurlow, Lazarus, Thompson & Morse, 2005). Furthermore, test results for students with disabilities were to be reported as a part of the total-population aggregate and again as a separate subgroup (Riddle, 2008; Thurlow, Lazarus, Thompson & Morse, 2005).

Perhaps the greatest impact of IDEA on educational assessment revolved around the measurement of students who, due to significant cognitive disabilities, were unable to participate in the general testing. Because all students were to be tested, IDEA required that an alternate assessment for this population must be developed (Lehr & Thurlow, 2003; Thurlow, Lazarus, Thompson & Morse, 2005), thus requiring all states to create such assessments. However, the 1997 IDEA did not specify the content or form of these assessments (Hager & Slocum, 2002), nor was this topic addressed in the 2004 reauthorized statute (DOE, Office of Special Education Programs, 2007). As a result, states have developed their own alternate assessments, and while there is no mandate regarding content, Lehr and Thurlow (2003) report that there has been a shift from “functional skills to student achievement of state standards.”

The form of the alternate assessments varies. Different states have developed different types of alternate assessments for their students with significant cognitive impairments. These alternate assessments include portfolios, checklists, rating scales, and sometimes even rely on analyses of students’ Individual Educational Plans (IEPs) as the alternate assessment (Hager & Slocum, 2002). In earlier years, the assessment of children with disabilities was regarded by most general-education teachers as a task only of concern to their special-education colleagues, but this is no longer the case. Rather, because many children who have special needs now are included in regular classrooms, today’s teachers are confronted far more frequently with the need to arrive at appropriate assessment approaches for children with disabilities. Clearly, these changes in the way special-needs children are currently assessed have been brought about by a series of landmark federal laws regarding how to educate and how to assess those children.

Looking back, then, at this brief overview of a half-century’s worth of federal legislation bearing on educational assessment, we have seen an ever-increasing impact of these federal laws on real-world assessment practices. Early versions of ESEA, for example, found educational testing used chiefly for monitoring whether federal funds for statute-specified populations were being effectively used for those populations. But in the early nineties, an important shift in purpose was seen when ESEA assessment results were to be used in making sure that all students, not just the disadvantaged students identified in Title I of the law, were achieving the curricular aspirations that had been set out for them. Moreover, the curricular aims for children with disabilities—albeit modified to some extent—increasingly became identical to the curricular aims sought for all students. This shift in orientation took place gingerly in IASA and in early versions of IDEA, but arrived with considerably more clout in the most recent renditions of both ESEA and IDEA.

At this moment, there seem to be few challenges to the belief that educational test results should be used to ascertain how well the nation’s students have been taught. Whether the federal government should play a prominent role in determining the nature and uses of U.S. accountability tests is, of course, a pivotal consideration in any rethinking of an appropriate federal role related to educational assessment.