IMPORTANT ISSUES OF EDUCATIONAL TESTING: LESSONS FROM THE NO CHILD LEFT BEHIND ACT

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Masters of Arts in the
Graduate School of The Ohio State University

By
Anne Christine Retz, B.A.

* * * * *

The Ohio State University
2007

Master’s Examination Committee:
Professor Ayres D’Costa, Advisor
Professor Dorinda Gallant

Approved by

[Signature]
Advisor
Graduate Program in Education
ABSTRACT

In January 2002 President George W. Bush signed the No Child Left Behind Act (NCLBA) into law and imposed the most expansive and regimented educational law in the history of the United States. Since the enactment of the Act and its extensive testing regime, controversies have arisen regarding the testing elements of the Act; but also who the Act may harm, as well as its fairness, both educationally and financially. The "unfunded mandates" debate has been popular in the media as well as in the legal system because of the staunch opposition to No Child Left Behind (NCLB) on the basis of the amount of money it will cost states and public school districts. Providing timely notice for supplemental services as well as the stipulation for highly qualified teachers in every classroom are two NCLB issues that have also garnered adequate attention in the courts.

The well-known and highly contested issue of adequate yearly progress (AYP), which must be demonstrated through test scores by every school and district for mathematics and reading has received the majority of attention during the previous five years of NCLB's life. The Department of Education has attempted to be more lenient in the rules and regulations initially outlined by the Act. This has spurred not only criticism from the educational community, but also praise from others. The issue of AYP is closely related to the massive and unprecedented level of student testing that has resulted from the enactment of NCLB. This unprecedented level of testing, unseen prior to
NCLB, has raised concern for the reliability and validity of the tests being used to measure student performance. With annually mandated testing occurring in grades three through eight, with a test in high school having been implemented in the 2005-2006 school year; this issue, and the possible numerous faults of the tests being used, will assuredly propel more scholarly research in the years to come as these tests are scrutinized more carefully.

This report is designed to discuss in detail the issues of the controversy around meeting AYP and the assessments used to measure it; legal issues involving NCLB; sanctions imposed on failing schools; the differing state academic standards, and some of the measurement issues related to the assessment measures of NCLB. This report aims to paint a balanced picture of what has happened in the past five years of NCLB’s life and to offer suggestions for new directions for improvement in what some believe is a comprehensive and worthwhile educational policy.
Dedicated to my loving and supportive parents, David and Christine, and my inspirational brother, Dave.
ACKNOWLEDGMENTS

I would like to thank my advisor, Ayres D’Costa, for his intellectual guidance and support throughout this adventure. His leadership and advice made this thesis possible and I am eternally grateful for all that he has done for me throughout this process.

I also wish to thank Dorinda Gallant who was especially helpful in this endeavor. Her direction and counseling during this process was instrumental in helping me to achieve my goals.
VITA

June 15, 1983................................. Born – New Brunswick, New Jersey

2005.............................................. B.A. Secondary Mathematics Education, The College of New Jersey

2005-2006................................. Mathematics Teacher, North Brunswick Township High School, North Brunswick, New Jersey

FIELDS OF STUDY

Major Field: Education
TABLE OF CONTENTS

Abstract............................................................................................... .ii

Dedication............................................................................................ .iv

Acknowledgements................................................................................... v

Vita..................................................................................................... vi

List of Tables........................................................................................ .ix

Chapters:

1. Introduction................................................................................... 1

2. Legal Issues................................................................................... 9
   2.1 The Case of Unfunded Mandates....................................................10
   2.2 Sanctions Imposed on Failing Schools and Highly Qualified Teachers...19
   2.3 Notice for Supplemental Services..................................................21
   2.4 High-Stakes Testing................................................................. 24
   2.5 NCLB v. Desegregation.............................................................. 29

3. Adequate Yearly Progress.................................................................33
   3.1 Numerous Changes to AYP and the Leniency of the DOE...............33
   3.2 Is 100% Proficient Achievable?....................................................40
   3.3 Sanctions Imposed on Failing Schools.........................................46

4. State Standards................................................................................49
   4.1 So What About the NAEP Test?..................................................53

5. Can One Test Really Provide So Much Information?.........................57
   5.1 Use of the Tests Generated by NCLB...........................................57
   5.2 Is all of this Test-Use Fair?....................................................... 62
   5.3 Measurement Issues................................................................. 64
   5.4 AYP Differences........................................................................68
6. Suggestions for Future Educational Policy Initiatives ........................................... 71

List of References ..................................................................................................... 79
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 NAEP National Average Test Scores-Mathematics</td>
<td>42</td>
</tr>
<tr>
<td>3.2 NAEP National Average Test Scores-Reading</td>
<td>43</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

In January 2002 President George W. Bush signed the No Child Left Behind Act (NCLBA or the Act) into law and imposed the most expansive and regimented educational law in the history of the United States. No Child Left Behind (NCLB) is itself a product of numerous other national educational policies and programs such as the Nation at Risk report of 1983 which led to a dramatic increase in state educational policy-making that had its roots in improving student achievement and most importantly setting academic standards for students to reach and educators to strive for. The measures taken after the Nation at Risk report led to further policies and programs such as Goals 2000, a program implemented by former President Bill Clinton, which stressed the creation of academic content standards, but did not impose specific regulations for carrying out the requirements outlined in the program (O’Neill, 2003). It is well-known that NCLB is actually the reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965, which was again reauthorized in 1994 as the Improving America’s Schools Act (IASA), which had previously been forgotten when other initiatives such as Goals 2000 came to the forefront. What is important about the ESEA and the well-known Individuals with Disabilities Act (IDEA) is that both spurred
significant increases in reporting, compliance, and administrative costs for states and local schools districts (Heise, 2006). McDonnell (2005) believes that IASA is the precursor that is most closely aligned with NCLB in terms of expectations for states and setting regulations and standards for states to meet. By examining these precursors to NCLB one can track the development of the policy and analyze what has improved and what has gone dramatically wrong. McDonnell (2005) argues that the evolution of the ESEA into the NCLBA “has changed the ways that at-risk students are targeted and served, and it has moved the federal role closer to the instructional core of local schools and classrooms” (p. 21). The NCLBA has three main requirements: wide-range testing of students in various grade levels, score reporting to determine proficiency levels for students, schools, and districts, and requiring highly trained professionals in the classroom. Within these three subdivisions there exist detailed rules and requirements that states, districts, and schools must adhere to in order to avoid sanctions imposed by the federal government.

The National Conference of State Legislatures (NCSL) outlined in a 2003 report to its state legislature’s presiding officers among others that some of the most crucial and wide-reaching requirements and regulations of NCLB are:

- All students in grades three through eight must be tested yearly in mathematics and reading as well as once in grades 10-12. Beginning in 2007, students must also be tested in science.
- Districts must collect data on individual students as well as report test data for the schools and districts. The states must publish the school,
district, and state data for public viewing by the beginning of the next school year as well as provide results to the Department of Education (DOE).

- Districts are required to report scores on subgroups on the basis of sex, race, socio economic status (SES), English language learner (ELL), and the special education population.

- States are required to provide assistance in disseminating scores for districts as well as provide assistance for schools that are designated for improvement after not meeting Adequate Yearly Progress (AYP).

- The funds received under the NCLBA from the federal government are designated to go directly to the local education associations (LEAs [better known as school districts]), but the state educational agencies (SEAs or state departments of education) are responsible for overseeing that the regulations as well as the penalties outlined by NCLB are being met.

- For a state to receive the appropriated funds from Congress under NCLB (a discussion on the language of the Act will be addressed later) it must first submit to the DOE a state plan for adhering to the NCLBA, which includes an “Accountability Workbook” outlining how the state will meet its self-determined AYP goal for each year leading up to 2014 when the state must reach a 100% proficiency standard for all students.
NCLB requires a highly qualified teacher in every classroom for the core subjects such as English, mathematics, science, and social studies. These teachers must be utilizing proven research-based instructional methods.

When schools or districts do not meet their AYP goals (various rules and restrictions of meeting AYP will be discussed later) they must notify parents in a timely manner and provide options for parents of the lowest performing and economically disadvantaged students, such as free tutoring services or allowing the students to attend higher-performing schools courtesy of the LEA (National Conference of State Legislatures [NCSL], 2003; also in Heise, 2006).

Fundamentally, NCLB was designed to raise standards. In Building on Results, a 2007 report released by the DOE, Secretary of Education Margaret Spellings discussed the theoretical background to NCLB, stating that “academic standards would be set by states, schools would be held accountable for results, and the federal government would support both with increased resources and flexibility” (p. 1); and that NCLB was designed to “steady academic gains until all students can read and do math at or above grade level closing for good the nation’s achievement gap between disadvantaged and minority students and their peers” (p. 1). This report was released at a time when deliberations were again coming to the forefront in the federal government about whether or not to rewrite and reauthorize NCLB for the remaining years of its proposed existence. It is designed to garner support not only politically, but also nationally among educators and parents. Spellings provides convincing data to support the DOE’s claim that NCLB
is “working, with test scores rising and achievement gaps narrowing” (U.S. Dept. of Ed., 2007, p. 1). The report asserts that total federal funding for NCLB rose 34% between 2001 and 2006 and that funding for Title I schools serving low-income students rose 45% in this same period (U.S. Dept. of Ed., 2007). Further, “states and school districts also have unprecedented flexibility in how they use federal funds in exchange for greater accountability for results” (U.S. Dept. of Ed., 2007, p. 1).

This exact issue of funding, and determining if states have truly received what they feel are the necessary funds to efficiently enact and adhere to NCLB, has created a bitter debate in the country that has infiltrated the courts and sparked scholarly debate. The funding debate, more commonly known as the “unfunded mandates” debate has produced a fight against the federal government for “upsetting the educational federalism status quo” (Heise, 2006, p. 127), and upset the legal and political standards that had previously been in place (Connecticut v. Spellings, 2006; Heise, 2006; Pontiac v. Spellings, 2005). This debate is one of many that has landed in the courts to be determined by judges after long-standing disputes between LEAs or SEAs and the DOE could not be resolved. The educational federalism status quo that Heise (2006) discusses in depth has created problems because “legally, NCLB safely navigates through existing constitutional requirements” (p. 151) but those working under the regulations of NCLB, particularly educators, feel that their LEAs or SEAs are not receiving the promised adequate funds and support from the federal government to meet the unprecedented requirements of the Act. Another closely related issue to the federal government’s power over states under NCLB deals with the lack of clarified guidance for the more minute
details. Moores (2005) states that "actual implementation of the law is inconsistent, due to the lack of clarity of some of its components and given the different ways in which various states and other jurisdictions are responding to the law" (p. 75). The unfunded mandates debate, as well as many others, have found their way into the judicial system and the interests of the federal government in the NCLBA have almost always proven to be upheld by the judiciary.

Each state has devised a unique plan for meeting the requirements of NCLB, particularly in meeting AYP each year so that by 2014, 100% of all students will be proficient in mathematics and reading. "Under the No Child Left Behind Act, schools must show incremental and linear progress toward the attainment of academic proficiency in 12 years. In addition, states must develop separate progress goals for subgroups of students" (Goertz & Duffy, 2005, p. 7). Since each state is attempting to meet its own goals it is difficult to assess true progress among states. For instance, State A might report that 45% of its eighth grade students are proficient in reading, and State B might report that 78% of its eighth grade students are proficient in reading, but if State B has a much lower proficiency standard than State A, it can meet AYP with much more ease. In addition to different proficiency standards for meeting AYP each year, states also have varying academic standards for their students that also contribute to making state-by-state comparisons difficult. The notion of using AYP to measure improvement, however, is valuable because as Goertz and Duffy (2005) assert, "reporting of assessment results allows the public to become aware of how a school and its students are achieving based on test scores and other data. The public can use this information to demand
improvement in their schools” (p. 6). These issues outline only a few of the debates related to AYP and state academic standards that have developed since NCLB was created.

Finally, there is the issue of the assessments or tests being used to measure improvement in America’s schools. Just as AYP predictions and state academic standards vary so significantly, so do the tests used to measure improvement and record AYP. Some states have much easier tests than others that has created a problem for comparisons to be made accurately between states. Heise (2006) notes the concern that allowing states to create their own proficiency marks for these tests will result in states diluting their proficiency standards. Jehlen (2003) believes that NCLB was designed to use the high-stakes assessments as pressure for schools to improve their own standards to match those of schools perceived to be better. O’Neill (2001) supports Jehlen’s idea and notes that “assessments are often a catalyst for educational reform” (para. 6). High-stakes testing has previously been noted to have high-stakes meaning for students, such as in the case of graduation tests or exit exams used in many high schools, but the tests of NCLB are high-stakes for the educators as opposed to the students. Wanker and Christie (2005) note how NCLB raises the stakes for states, districts, and states when they fail to demonstrate academic improvement. In 2003, Zirkel proclaimed that “high-stakes testing, which has already been the subject of a long line of court decisions, is bound to be subjected to increased litigation in the wake of the No Child Left Behind Act” (p. 712). This paper will also examine just how testing and the courts have coexisted to enforce the policies of NCLB, instead of breaking them down.
This report is designed to discuss in detail the aforementioned issues of the controversy around meeting AYP and the assessments used to measure AYP; legal issues involving NCLB; sanctions imposed on failing schools; the differing state academic standards, and some of the measurement issues related to the assessment measures of NCLB. This report aims to paint a balanced picture of what has happened in the past five years of NCLB’s life and to offer suggestions for new directions for improvement in what some believe is a comprehensive and worthwhile educational policy.
CHAPTER 2

LEGAL ISSUES

As far as the legal system has been concerned, most of the issues that have found their way into the higher courts involving ramifications of NCLB have centered on unfunded mandates, lack of highly qualified teachers, giving notice for supplemental services, and high-stakes testing. Each issue has had an uncanny focus on the language of the NCLBA, which has required a fine-tooth comb look at the wording of specific sections of the Act. One will see how the history of educational federalism and the legal system have worked together to generally substantiate the regulations and requirements of NCLB. The NCLBA was written brilliantly and survives on the fact that throughout history the Supreme Court has mainly removed itself from educational matters because we have elected to put education in local hands (Heise, 2006). Furthermore, NCLB does not force any overtly ludicrous dictates on states (Heise, 2006; Moores, 2005). "The United States Supreme Court has repeatedly held that public education is 'perhaps the most important function of state and local governments'" (San Antonio Independent School District v. Rodriguez, 411 U.S. 1, 29 (1973) in NCSL, 2003). In our country’s history the federal government’s capacity in education was limited to providing funding for disadvantaged groups such as the economically disadvantaged and persons with
disabilities. "The Supreme Court has been reluctant to interfere with the states' rights to oversee education and usually will do so only when necessary to protect freedoms and privileges guaranteed by the United States Constitution" (Buchman et al., 2000, p. 3-2).

Heise (2006) examined the widely held belief that local school boards are "comparatively better positioned to set education policy in a manner that reflects local conditions and preferences" (p. 131). Finally, Heise (2006) states:

> Even NCLB's harshest critics must applaud the strategic genius it embodies: an elegant use of political-economic leverage that generates policy coercion upon states that extends far beyond the reach of NCLB funds. By astutely targeting one critical link (student assessment) in the tightly woven educational policy chain and understanding the inexorable tether that binds that student achievement variable to a host of other distinct, though related, policy variables, NCLB vividly illustrates the high art of policy leverage...Thus, through NCLB the federal government can achieve its policy goals on the proverbial financial backs of states and local school districts (p. 141).

*The Case of Unfunded Mandates*

There are three distinct court cases that have explored the legal ramifications of NCLB in a financial sense. Of these three cases, the two most publicized are *Pontiac v. Spellings* and *Connecticut v. Spellings*. In *Pontiac v. Spellings*, Secretary of Education Margaret Spellings argued that the "court lacks subject matter jurisdiction because the plaintiffs lack standing, and that the issue, 20 U.S.C. § 7907(a), does not provide the relief the plaintiffs seek" (*Pontiac v. Spellings*, 2005, para. 14). This helped foster the
case's eventual dismissal on behalf of Spellings. In this case the plaintiffs included several school districts in the states of Michigan, Texas, and Vermont, as well as the National Education Association (NEA) and various NEA affiliates. Together the plaintiffs alleged that the NCLBA was imposing unfunded mandates, or directives for SEAs and LEAs to follow that were not being adequately funded by the federal government. Furthermore, the plaintiffs suggested that these unfunded mandates were specifically forbidden by clauses within NCLB. The clause in question, commonly known as the Unfunded Mandates Provision, 20 U.S.C. § 7907(a) of the “Prohibitions on Federal Government and use of Federal funds” section of the NCLBA, states that as general prohibition:

Nothing in this chapter shall be construed to authorize an officer or employee of the federal Government to mandate, direct, or control a State, local education agency, or school’s curriculum, program of instruction, or allocation of State or local resource, or mandate a State or any subdivision thereof to spend any funds or incur any costs not paid for under this chapter [italics added] (NCLBA, 2002).

The complaint was dismissed due to the language of the NCLBA, which generally states that through the Act “Congress clearly meant to prohibit federal officers and employees from imposing additional, unfunded requirements, beyond those provided for in the statute” (Pontiac v. Spellings, 2005, p. 5). In the case summary, it states that the NCLBA “cannot reasonably be interpreted to prohibit Congress itself from offering federal funds on the condition that the States and school districts comply with the many statutory requirements, such as devising and administering tests, improving test scores,

---

1 For further detail of this case see the case summary and commentary (Pontiac v. Spellings, 2005).
and training teachers” (Pontiac v. Spellings, 2005, p. 5). The NCSL (2003) outlined the unfunded mandates case in terms of Section 9527(a) (Prohibitions on Federal Government and use of Federal Funds) of the Act and the plain meaning of the section which many involved believe to mean that “states, or local subdivisions, do not have to spend funds on the costs of NCLB that are not paid for by the Act itself” (Section 3B, para. 1) and further how the NCSL believed a state or LEA could make a case against the federal government noting that the aforementioned section has the potential for federal court challenges. As of yet though, no state or educational agency has been successful in pursuing a case against the federal government under the guise of unfunded mandates.

Heise (2006) has contributed significant work to the issue of unfunded mandates in NCLB and education federalism. Heise (2006) makes the case that NCLB “constitutes a permissible exercise of Congress’s conditional spending authority under the Court’s present interpretation of South Dakota v. Dole” (p. 128) Under the NCLBA, Congress is authorized to spend a significant amount of money on education, but this does not translate to every dollar of the authorized amount to be spent. Congress agrees to provide financial resources that can be restricted by endorsed government agencies if rules and regulations are not adhered to properly in the predetermined manner. Heise (2006) also contends that the public must remember that NCLB is not a mandate because it is a program for which participation is voluntary. He also makes an astounding point that states lose their credibility in the case of NCLB as an unfunded mandate because “at some point the 100% participation level in a voluntary program begins to erode confidence in assertions that NCLB costs states millions—if not billions—of dollars,”
noting that if the financial damage was truly so severe as they claim they would not participate (Heise, 2006, p. 149). Finally, Heise (2006) does note the importance that if states feel so strongly compelled to participate in NCLB, a voluntary program that imposes such significant financial costs, then the coercive force of the federal government is much more intense than perceived.

The *Connecticut v. Spellings* case brought to the forefront many different issues of the NCLBA, but the interest of this investigation lies in the unfunded mandates portion. The case is centered on the responsibilities of the State and of Secretary Spellings under the Act. It is important to note that Connecticut did not challenge the Act itself, but rather Spellings and whether or not she was in full compliance with the Act (*Connecticut v. Spellings*, 2006). Connecticut was seeking clarification of the Unfunded Mandates Provision (20 U.S.C. §7907(a) as stated earlier) and argued that Spellings’ interpretation was incorrect. Spellings argued that her interpretation was indeed in compliance with the Act and further challenged the Court’s general jurisdiction in the case (a general challenge seen in many of these cases). In the court discussion of this section of the Act, Connecticut interpreted the section to imply that the provision must supply full federal government funding for different provisions of the Act, and that in addition, if the necessary funds are not allocated then the Secretary of Education is allowed to release a state from compliance with those provisions of the Act that cannot be financially met (*Connecticut v. Spellings*, 2006). In Spellings’ response to this, she argues that the Unfunded Mandates Provision is null and void in this case because it

2 For a further discussion on jurisdiction, particularly subject matter jurisdiction, see legal challenges in this section (*Connecticut v. Spellings*, 2006; or *Pontiac v. Spellings*, 2005).
applies to provisions made by federal officials, but the NCLB was imposed by Congress. This is the same argument made in the *Pontiac v. Spellings* case (*Connecticut v. Spellings*, 2006). The two counts of the State’s claim against Spellings that explored the interpretation of the Unfunded Mandates Provision—the Spending Clause and Tenth Amendment that dealt with the financial matters of the case—were dismissed because the Court determined that it lacked subject matter jurisdiction in these matters.

The last (and lesser known) case involving inadequate funding under NCLB is the *Save Our Schools v. D.C. Board of Education* (2006), in which the plaintiffs were suing the District of Columbia’s Board of Education, among other parties such as Spellings; over the inadequate, unfair, and purposeful lack of funding being provided to the lowest performing public schools in the District of Columbia. In this case, the plaintiffs argued that the funds received from the federal government were being misappropriated. The counts of the complaint that were directed toward Spellings regarding the NCLBA and involving purposeful inadequate funding under the provisions of NCLB were dismissed for independent reasons. In plain fact there was no case against Spellings because the funds provided by the federal government were re-directed for states to allocate just as the Act permits and directs. In this specific case, the funds were re-directed away from underperforming schools but the court did not consider this to be conjectural injury (*Save Our Schools v. D.C.*, 2006). The other claim against Spellings in this case stated that

---

3 The “Spending Clause” of the United States Constitution permits Congress to “lay and collect Taxes, Duties, Imposts, and Excises, to pay the Debts and provide for the common Defence and general Welfare of the United States” (Article I Section 8, Clause 1 of the U.S. Constitution as cited in *Connecticut v. Spellings*) and that “Incident to this power, Congress may attach conditions on the receipt of federal funds, and has repeatedly employed the power to further broad policy objectives by conditioning receipt of federal moneys upon compliance by the recipient with federal statutory and administrative directives” (in *South Dakota v. Dole* as referenced in *Connecticut v. Spellings*, 2006). Through the Tenth Amendment, Congress can withhold federal funds if a State does not comply with national laws.
Spellings, under NCLB, was liable for a violation of the Fifth Amendment. This was dismissed because the claim fell under state law, and it was determined that Spellings was acting under federal law and was therefore null and void.

Each of these cases reflected the matter of funding under NCLB. This issue is perhaps the most hotly debated legal issue in reference to NCLB. Hoff and Richard (2004) cite the Ohio Department of Education, which had released an earlier study that stated it would spend $1.5 billion to meet the provisions of NCLB. This was twice as much as Ohio was then receiving in federal funding for education. Early in 2004, Virginia’s House of Delegates passed a resolution stating that NCLB would cost Virginia millions of dollars that it simply did not have (Hoff & Richard, 2004; Welner & Weitzman, 2005). In addition to Ohio and Virginia, other states that have been vocal in expressing their distress about the perceived unfunded mandates of NCLB include North Dakota, Utah, Hawaii, Louisiana, Minnesota, Nebraska, and Vermont (Hoff & Richard, 2004; Keller & Blair, 2003; Mathis, 2005; Welner & Weitzman, 2005). New Hampshire’s superintendents claimed that the state should reject Title I funds because they would lose less money than they would if the state chose to comply with the stipulations of the Act. Three suburban towns in Connecticut opted out of participating in Title I programs (Méndez, 2003 in McDermott & Jensen, 2005). In response to these and similar outcries of other states or interested organizations, the Bush administration retorts that these states and organizations continue to overlook one important fact: prior to NCLB the states had already been committed to raising student achievement, and
therefore should already be expecting to spend money to achieve this goal (Hoff & Richard, 2004).

In 2005, Utah passed a bill that put Utah's educational policy ahead of that of the federal NCLB law (McDermott & Jensen, 2005; Sack, Robelen, & Davis, 2005). The discrepancy between Utah's state education testing system, called U-PASS, and the federal mandate lies in the reporting of test scores. U-PASS uses a different system and has different requirements for reporting. To change the system to adhere to the requirements set forth by NCLB; large quantities of financial resources would need to be used (Sack, Robelen, & Davis, 2005). After a long political debate, Utah decided to opt out of the law and not receive federal funds to support its Title I schools—to date an unprecedented move (Mathis, 2005).

In 2004 the NEA released a report stating that in the previous fiscal year states received only $18.6 billion of the $26.8 billion in federal funds that were authorized under the law (Hoff & Richard, 2004). The federal government has argued that "authorized" spending does not equal funds that will be fully allocated, whereas the states and other interested parties frequently argue that because the money is authorized by Congress to be spent it should be distributed for use, especially if deemed necessary to achieve academic goals within states. Mathis (2005) has done extensive work analyzing the unfunded issues debate between the federal government and state and local governments. As an example of how the notion of "authorized spending" can be interpreted differently by different parties he cites the 2003 Secretary of Education, Rod Paige, who noted a 51% increase in federal appropriations versus the various state
legislators who argue that the federal government at best provides 8.2% of total spending on education. There is a vast difference between 51% and 8.2% as Mathis notes, and clearly any interested party can find a way to make data and statistics work in their favor to support their position. Mathis (2005) explores six different definitions of the term “funding,” each being the backbone of a different party’s position in this fulsome debate of NCLB as an unfunded mandate.

The six definitions offered by Mathis (2005) are:

- Relative funding—The percentage increase in the federal appropriation
- “Money left on the table”—Unexpected federal monies at the end of the fiscal year
- Authorization level—What the law said would be appropriated
- Law’s definition of full funding—Dollar cost of funding all eligible students at legally prescribed level
- Administrative hard costs—Estimates limited to cost of administering new features of the law
- Costs of teaching children to standards—Based on adequacy studies of what it costs to bring children to standards (p. 94)

Mathis (2005) provides a more elaborate discussion on each definition and provides illustrations and comments for each. For example, when he discusses the third definition as authorization level he notes that President Bush asked for $13.3 billion for Title I funding for the 2005 fiscal year, but the authorized level under NCLB was $20.5 billion. Rod Paige, the previous Secretary of Education, notes that authorization levels
do not equal financial commitments by the federal government. Under this long-held mandate that authorization does not equal allocation states can no longer continue to attempt to make arguments, for they are only wasting precious time that could be spent finding ways to reallocate state and local funds to help fund educational endeavors. This discussion of authorization levels is the most highly publicized and most widely used in the media, legal, and political debates (Mathis, 2005). McDermott and Jensen (2005) have noted that statutes involving conditional spending, such as that seen in NCLB, “routinely survive legal challenge” because “the judiciary’s logic is that the conditions attached to federal monies do not compel behavior by states or individuals, but merely ‘encourage’ states and individuals to ‘conform to federal policy’” (citing Justice Sandra Day O’Connor in U.S. v. Lopez, 1995, p. 41). The issue of whether or not NCLB is an unfunded mandate will never truly be solved in the public’s eye until the courts, politicians, educators, and the public all agree on a single definition for funding. Legally, the courts have the final say, and as of now, they have continued to side with the federal government, and that authorization does not equal financial commitment.

Other arguments in this unfunded mandates debate include Hoff and Richard (2004) who stress the bipartisan complaint arguing for the cause of unfunded mandates. This is not beneficial to the current Bush administration. Hoff and Richard (2004) cite Eric A. Hanushek from Stanford University who directs attention to what he believes is the real issue: that the state and federal governments do not know how to separate the added cost of NCLB from the monies that would be spent for continuing efforts to improve schools in any case. It raises the question of whether or not states and local
school districts were given enough time to reorganize in an efficient manner to ensure that there is no frivolous spending and to find ways to use all the provided funds—federal, state, and local alike—to design the best budgets and programs for each district. Moores (2005) has added to this argument that Congress authorized to support 40% of the excess cost of educating and providing special services to students with disabilities, but that no more than 20% has ever been supplied. But Welner and Weitzman (2005) point out that “Congress is not legally bound to match the amount of earlier authorization with the later appropriation” (p. 242). This has only further supported the federal government’s case in the aforementioned legal complaints regarding unfunded mandates.

Finally, in this debate of unfunded mandates, Heise (2006) argues that this issue should provoke society to question that as a policy “is it prudent to permit the federal government to exercise critical education policy influence beyond the extent of its financial contribution to states and local school districts?” (p. 129). He also notes that the constitutional and legal texts say little, if anything, about policy-making authority (Heise, 2006). He notes too that NCLB is a voluntary program for states, but if a state wishes to receive financial education aid it must participate and follow the set rules and regulations (Heise, 2006). He closes his argument by providing a possible solution, stating that “educational policy control should presumptively fall to the level of government willing to shoulder the relevant costs” (Heise, 2006, p. 153).

Sanctions Imposed on Failing Schools and the Issue of Highly Qualified Teachers

There have not been many highly publicized court cases that have dealt with the issue of highly qualified teachers or sanctions imposed on failing schools, and neither
issue has been nearly as popular or well-known as the unfunded mandates debate. In 2003 a federal judge in Kansas rejected the lawsuit brought by Brian K. Kegerreis who argued that the NCLBA is unfair and unconstitutional because it seeks to hold only school personnel accountable if students do not meet the specified proficiency standards on new state tests. The judge noted that the United States government is generally protected from such lawsuits unless it has expressly waived its sovereign immunity (Kegerreis v. USA, 2003). Furthermore, when the case was argued Kegerreis could not make a claim because he had no current injury inflicted by the NCLBA and that the circumstances that could inflict injury (in this case schools not meeting AYP and teachers possibly losing their jobs) are not only hypothetical but would occur over the next 11 years (Kegerreis v. USA, 2003; Robelen, 2003). In 2003, while the NEA was looking for allies for their case against the federal government regarding funding of the new ESEA law (now known as NCLB), it pleaded for changes to be made to the legislation including changing the definition of “highly qualified” for teachers and asking for less proof of academic-content knowledge. This change of definition was sought because the NEA felt that the federal government was not providing enough money to aid states in complying with this provision of the federal law and therefore not enough federal aid was being supplied to train teachers more effectively (Keller & Blair, 2003). As Moores (2005) argues, “as with the inadequacies related to statewide testing, there is no mechanism to ensure consistency across states in criteria for highly qualified teachers” (p. 79), this point is well noted and it should perhaps be a point for further consideration for the future of NCLB.
Finally, in California, the state board of education was sued by both the Californians for Justice and the California Association for Community Organizations for Reform Now over the board’s definition of “highly qualified” teachers, contending that the definition being used included emergency-certified teachers who did not meet the requirements of NCLB (Walsh & Sack, 2003). This case has no published outcome and the definition has since been changed by California to meet NCLB standards for highly qualified teachers.

Notice for Supplemental Services

There have been two cases that have generated discussion on behalf of notifying parents about supplemental services for students of schools that are not meeting set proficiency standards. In these cases, the NCLBA itself was not questioned, but state and local education agencies were being questioned with regards to their compliance with the sanctions of NCLB. In the case of the Association of Community Organizations for Reform Now v. New York City Department of Education (Ass’n of Cmty. Orgs. for Reform Now v. New York City Dep’t of Educ.) two sets of parents with children in New York state Title I school districts that were designated as low performing filed suit against the New York City Department of Education for violating NCLB. The parents partnered with the Reform Now group and filed a class action federal lawsuit for alleged violations of NCLB and the New York state constitution (Zirkel, 2004). The parents argued that they were not given sufficient or timely information about resources for parents to sign up to change schools or take part in other supplemental services. Under NCLB if a school is marked as low performing for two years in a row after not meeting
AYP in some manner, then “schools must offer their students the options of supplemental services and transfer” (Zirkel, 2004, p. 479). Zirkel (2004) continues, “the supplemental services options applies only to children from low-income families, but must be made available only so long as there are sufficient funds allocated for that purpose, and may be waived, upon district request, by the state” (p. 479). He further notes that districts must provide notice to parents when schools or districts themselves are placed in the school improvement, corrective action, or restructuring category (Zirkel, 2004). The plaintiffs argued that information was provided to some parents and not others who were equally entitled to receive the information on school transfers or other services. They also argued that when the parents attempted to gather information they suffered unnecessary hardship in doing so. In June 2003 the federal district court dismissed the case, claiming that the plaintiffs had no right to sue the Department of Education and that “Congress did not intend to create individually enforceable rights with respect to the notice, transfer or [supplemental educational services] provisions contained in the NCLBA” (Zirkel, 2004, p. 480). Within his close analysis of the case, Zirkel outlines the courts decision in three parts. He states that “first the court concluded that the notice, transfer, and supplemental services provisions do not contain ‘rights-creating’ language focused on an individual or a class of individuals” (Zirkel, 2004, p. 480). He then notes that “second, these statutory provisions have only an ‘aggregate focus’ rather than an individual entitlement, as evidenced by their limitations and exceptions” (Zirkel 2004, p. 480). Zirkel (2004) says that the Act does not “include an administrative or a judicial procedure for individual use, providing the secretary of education with the sole enforcement authority and then only
directly against states, not even districts” (p. 480). Finally, he discusses the language of the Act, and just as one sees in the unfunded mandates debates, the language provides loopholes for the federal government. Zirkel (2004) says “the Act has a centralized enforcement mechanism, evincing a congressional intent ‘to avoid the possibility of individual lawsuits and multiple interpretations’” (p. 480). He claims that because of this, litigation against NCLB or perhaps in relation to NCLB provisions and the actions of the federal government under NCLB, has little or no chance to succeed (Zirkel, 2004). When NCLB was written, lawmakers and policymakers had history on their side because the Supreme Court has traditionally backed away from cases that deal with educational issues because education has always been in the hands of state and local governments. The federal courts as well as the Supreme Court, in particular, only get involved in educational matters if they feel a severe strike against the constitution has been made or if an issue has been improperly handled by the lower courts (Zirkel, 2004). In their analysis of this same case, Walsh and Sack (2003) said that this case is most important for raising the question about whether or not plaintiffs have the right to sue school districts of states that are not properly complying with NCLB. The outcome of this case has put the issue of suing on individual rights to rest.

The second case, that of Fresh Start v. Toledo Board of Education (2005), involved the Fresh Start Academy, a provider of supplemental services such as tutoring, suing the Toledo Board of Education (Toledo BOE) stating that the Toledo BOE purposely blocked Fresh Start and other groups offering similar services from receiving funds, that the Toledo BOE misappropriated funds, and gave preferential treatment to
others and thereby inflicted financial harm on Fresh Start. When the Toledo schools were indicated as low performing, Fresh Start wished to provide tutoring services to Toledo’s students; however, the Toledo BOE eventually chose not to use Fresh Start despite its believed fulfillment of requirements to be deemed an acceptable provider of supplemental services. The motion was dismissed because “Fresh Start’s claim depends upon rights allegedly conferred by the NCLBA, but that the Act does not confer any private rights upon Fresh Start” (Fresh Start v. Toledo BOE, 2005, para. 10). These two cases did not directly challenge the NCLBA in any way, but rather they showed the strength of the law in requiring SEAs and LEAs to provide supplemental educational services when schools or districts are designated as low performing; and that parents are well aware of their rights and are committed to providing the best education for their children that they can. These cases display the legal strength of NCLB that is emitted through its careful legal language and sole imposition on states and their responsibilities to education rather than LEAs, educators, or parents.

High-Stakes Testing

Historically, high-stakes testing, which generally includes standardized tests and assessments, were used to make important decisions for individual students and have been known to carry heavy consequences (Miksch, 2003). The earliest challenge to high-stakes testing began with the Debra P. v. Turlington case in which the federal courts set legal standards for exit exams by addressing three major issues: “(1) curricular/instructional test validity as a substantive due process requirement; (2) prior notice required to satisfy procedural due process; and (3) vestiges of segregation as intent
to discriminate under the equal protection clause” (Phillips, 1993, p. 11). Most lawsuits against high-stakes testing have been with regards to due process or regarding the issue of students not having enough notice and therefore time needed to pass the test and being provided with the chance to be adequately prepared (Miksch, 2003; O’Neill, 2003). And there have been challenges to high-stakes testing regarding adverse impact on minority students and that tests have a “discriminatory effect” on these minority students (Miksch, 2003, p. 54). In the Rene v. Reed case, the lawsuit was brought on the basis of special needs students who were not allowed the proper modifications in order to take a test, in addition to the plaintiffs’ claims that the special needs students had not been equally or adequately prepared for the material on the test in comparison to their regular-education counterparts (Miksch, 2003, O’Neill, 2003).

What is special about the NCLBA is that the assessments used are not particularly high-stakes for students, but rather for the teachers, as the Kegerreis v. USA case tried to prove through the fact that teachers are the ones who will suffer by the possible loss of a job if students continue to not improve at the prescribed rates. O’Neill (2003) who is well known for his research on legal aspects of high-stakes testing, notes that “NCLB does not require that high stakes be imposed on students (nor does it forbid this), but it does hold the system’s feet to fire where students fail to show adequate yearly progress” (p. 633). O’Neill (2003) has contributed much work on the high-stakes of high schools’ exit exams and notes that states that already use these high-stakes exit exams will most likely use them to meet their NCLB high school testing requirement so “most state promotion exams will be linked into the NCLB framework and subject to its reporting
and accountability requirements as well” (p. 634). He continues that as the NCLB testing regime continues, the requirements imposed by NCLB may turn the tide of states’ use of high school exit exams in the face of pressure to meet AYP (O’Neill, 2003). With the enormous research base on high-stakes testing and the legal aspects and challenges that have resulted in the use of such tests, it might only be a matter of time until the rules under NCLB require states to have more rigorous laws that are subject to public scrutiny and litigation (O’Neill, 2001, 2003). Finally, O’Neill (2003) notes that the mandatory federal testing requirements of NCLB will force boycotters of high-stakes testing to “either give up the fights or redouble their efforts” (p. 659). High-stakes assessments have been greatly improved over time, but with the increased number of tests that test publishers have needed to generate, one should question if these tests are satisfactorily holding up to the aforementioned legal standards for tests.

Prior to NCLB, exit exams were popular to address student proficiency, but also as a means to allow teachers to identify student weaknesses and work to improve them. Miksch (2003) points out that the federal government’s theory in using testing as the main proponent of NCLB is “the hope is that exams will motivate students to work harder and allow teachers to identify and address student weaknesses” (p. 53). Clearly the federal government wanted to make sure that every state was participating in this effort, but that it occurs earlier in and throughout schooling rather than the previous trend of using such assessments at the end of high school. The government appears to have taken the educational theory that was developed with the original high-stakes tests and applied it to a much larger scope of individuals that now includes not only students, but
also teachers and administrators who are now held accountable for student deficiencies. As discussed later, the research varies in results to determine if high-stakes testing pushes teachers to work harder and students to achieve higher results. A study done by researchers at Arizona State University states that with all of the testing done over the past 20 years, there has been no discernable trend of improvement or decrease in scores. However, they did find that schools whose scores improved dramatically were spending an enormous amount of time teaching to the test (Jehlen, 2003). However, Lewis (2000) notes that many people, particularly politicians, feel that high-stakes testing achieves its purpose in pushing both students and teachers because the data they use shows an increase in test scores.

In the Arizona v. United States Department of Education (USDE) case, which was only decided in February 2007, the Arizona Department of Education (ADE) was looking for declaratory relief concerning a specific section of NCLB that declares that a school or district can appeal an identification of improvement, corrective action, or restructuring if it believes that there is an error in this identification based on a statistical error or an equally substantive reason (in 20 U.S.C. § 6316(b)(2)(B)), a provision of the NCLBA. The ADE had been granting appeals by schools who were failing to meet AYP due to a change in the Arizona testing policy that now forced all limited English proficient (LEP) students to take the standardized tests in English, which is in compliance with NCLB. Upon review, the USDE and Spellings found ADE in violation of the actions permitted by NCLB. This case found its way into the U.S. District Court of Arizona, and like many of the other cases that contested the meaning of a provision of NCLB, was dismissed for
a lack of subject matter jurisdiction. The court deemed that Spellings’ actions of reversing the rulings of the ADE were in compliance with the permissions of the NCLBA, and that her actions were under the guidance of her specialty (Arizona v. USDE, 2007).

Finally, there is the Reading School District v. Department of Education in which the Reading school district had several of its schools placed on the Warning list or on the in Need of Improvement list (dependent on Title I status) for failing to meet AYP two years in a row. The district felt that its schools were unfairly placed on these lists and wanted the court to determine if the DOE provided adequate technical assistance to the Reading school district based on the provisions on NCLB, if the failure of the DOE to provide native-language testing was a discriminatory act, and if the “N” number (number of people in a subgroup needed to determine whether the subgroup will be designated as a group that must meet AYP individually based on statistical methods) was indeed based on sound statistical methods. The district felt that it was forced to spend money it did not have to meet the requirements of NCLB and that it simply had no more funds to designate to the school improvement plan that was deemed necessary. The court not only decided that the technical assistance issue was not relative to the plaintiff’s case, but that even if it was, NCLB does not mandate that all technical assistance must be provided at one time; and that rather, NCLB stipulates that technical assistance should be provided throughout the improvement plan. As for the native-language testing discrepancy, the Act itself does not make the provision to supply native-language tests mandatory, and that Spellings was correct in her decision that at the time it was not feasible for
implementation of such tests in an accurate manner, so this expectation could not be held by districts. Finally, for the “N” number debate the court disagreed with the plaintiffs’ argument and felt that the DOE had used sound statistical methods, and recognized that at no time throughout the process of hearings and appeals prior to the current case did the district offer a more fitting “N” number or contest the DOE’s decision as it was being made. The court notes that NCLB is not a regulation of the DOE, but that the DOE and Spellings have just reason to impose and interpret regulations as they see fit because NCLB falls within the department’s sphere and area of expertise (Reading v. DOE, 2004).

_NCLB v. Desegregation_

An interesting dilemma occurred in 2004 in Pinellas County, Florida. In Pinellas County there was an existing court order that schools must be desegregated in order to promote a well-rounded educational environment for all students⁴. With the enactment of NCLB and the provision of school choice for low-performing students with a low SES, a conflict arose between the court-ordered desegregation and the option to attend a better school. This case presented a sizeable conflict between a long-held federal court ruling and the federal statute of NCLB, which the DOE felt should be determined to be more important in this case. NCLB mandates that if a school is designated as needing improvement the district must not only inform parents, but give priority preference to low-income parents who choose to send their children to a better performing school within the district. This stipulation affected the desegregation order that was already in effect, and by following protocol laid out in the desegregation plan and NCLB the

---

⁴ See DeBray, 2005 for more extensive background information on the desegregation court order.
districts needed to consult with a federal judge to determine how to comply with both NCLB and the desegregation order, and which statute takes precedence over the other. There were many schools that were marked as needing improvement and the district was looking to reduce class sizes all over the district to enhance learning for the overall population of students. When the need to offer an opportunity for change in schooling to low-income, low-performing students arose, the district simply did not have enough space to place all of the low-income students in the better-performing schools without displacing the desegregation law that was designed to equalize the schools racially. The only way in which the district could have followed the requirements for school choice as outlined by NCLB, would have been to do away with the desegregation order, a order that had taken many years to effectively devise and implement, that had been in place for quite some time, and that was valued by the community. DeBray (2005) notes that this case “highlights the problem of local consent and bargaining in the implementation of federal education policy” (p. 172). School districts and states have worked hard to meet the standards of NCLB and this case is an example of ambiguous choices in which no defined answers are in place. This in turn has created a lawsuit that costs the public more money and puts education on hold for those trying to find a better education elsewhere. When the case was decided, the school district won and was able to continue its desegregation plan, which they perceived to be more important to keeping their schools equal and effective, as long as it was complying with the stipulations of NCLB as best it could; and most importantly, the federal government could not decrease or take away its funding from the school district because it was not in full compliance with the terms set
forth by NCLB. This is a prime example of how the federal government and local
governments can work together to achieve harmony in the schools, and most importantly
for the students.

As noted, there are many different instances in which NCLB has found its way
into courtrooms on various levels, but no case has truly been successful in displacing
NCLB from its upper echelon position in federal law. The unfunded mandates debate
will undoubtedly continue until the next federal education law comes through, and even
then, any interested party can make a case for its position on funding based on Mathis’s
(2005) six definitions of funding. Zirkel (2004) argues that he believes that the only
opposition to NCLB that has any real chance in the courts is “administrative criticism of
insufficient funding, and individual pleas on behalf of some segments on the special
education community” (p. 480). But we have seen two cases that specifically identified
the perceived unfunded mandates of NCLB that have failed to increase federal funding
for states; and as of yet, there have been no notable cases involving the special education
community, and the position of the individual pleas has failed in the courts. Further, as
will be discussed later, the DOE has been making adjustments since the enactment of
NCLB, particularly with regards to the special education population and the necessary
testing requirements for this particular population. The language of the bill has made it
unlikely that any party will successfully challenge NCLB in the courts and the NCLBA
has safely maneuvered through political and public criticism both in and out of the court
system. The debate on supplemental services and adequately providing these services
with notice has also been a significant battle in the courts. In 2003, Miksch noted that
“the No Child Left Behind Act implies the promise that all students become academically proficient. This may lead to students and parents being able to sue a school because a child has not received an appropriate education” (p. 57) but we have seen that the right to sue as an individual under NCLB is null and void and will lead to a decision in favor of the federal government. Zirkel (2004) determines that “California’s allegedly inadequate implementation of the Act’s ‘highly qualified teacher’ requirements and our New York case, following the Supreme Court’s limiting standards in Gonzaga, seems to suggest that such litigation faces a dead end” (p. 480). Zirkel seems to have predicted an accurate future of such litigation that the NCLBA does not protect the rights of, nor provide rights for, individuals. Zirkel (2004) recommends that “the route to revise or reverse NCLB is via the political, not the judicial process” (p. 480). However, Keller and Blair (2003) note that using the political route to modify NCLB could cause serious conflict. Republican-led states could suffer consequences for opposing their Republican president, and Democratic-led states are also in an uncomfortable position because of the strong bipartisan support for NCLB when it was enacted. Regardless of what transpires in the political arena, it has been duly noted that in the past five years little or no legal progress has been made in changing any of the disputable requirements or mandates of NCLB.

---

1 The Gonzaga v. Doe case set precedent that individuals may not bring a civil rights suit to obtain compliance with the requirements of a piece of federal legislation, in this case the Family Educational Rights and Privacy Act (Zirkel, 2004).
CHAPTER 3

ADEQUATE YEARLY PROGRESS

The Numerous Changes to NCLB's AYP Provisions and the Leniency of the DOE

Throughout the existence of the NCLBA as a federal law there have been numerous changes to some of its requirements, particularly AYP, which is one of the most debated issues of the law and which also causes the most conflict. AYP is used to determine if students and schools are improving each academic year. Each district must set AYP goals for each school as well as the district overall and must reach 100% proficiency in reading and mathematics by the year 2014. After the initial AYP cycle many states “shifted a portion of their investments to the study of that first NCLB AYP cycle: examining what worked and what required refinement” (Forte Fast & Erpenbach, 2004, p. 1) and this led to many of the modifications that states began to ask for in 2003 and 2004. One area under AYP that has undergone significant changes is that of AYP predictions for the special education population. In June 2003, the DOE was allowing schools to use out-of-level assessments for students with disabilities, which allows students with disabilities the opportunity to take a test that meets NCLB testing requirements but that is also aligned to the needs and instructional level of the student (Forte Fast & Erpenbach, 2004). Another important AYP change that occurred was that
of allowing states to use standard rounding rules for calculating AYP, so ideally for some scores that were just on the border of meeting AYP predictions, these scores could now count as passing and schools and districts could meet AYP for a given year. Many states have been approved for use of a form of statistical estimating known as confidence intervals that allow for a small degree of error in score reporting. States have been granted permission to use 95% or 99% confidence intervals for making AYP predictions. This means that a state can be 95% or 99% positive that the true AYP score is in the given range (Center on Education Policy [CEP], 2007; Forte Fast & Erpenbach, 2004). The use of confidence intervals, especially 95% or 99% confidence intervals, allows a larger range for student test scores to fall into. These large confidence intervals have serious implications. Using the larger confidence intervals makes it easier for schools to meet their AYP predictions by using less stringent criteria for test scores to meet.

In the DOE’s 2007 report, *Building on Results*, it noted a continued commitment to promoting further flexibility for states and local school districts, so it proposed:

Additional flexibility, options and tools will be offered to help states improve their accountability systems and demonstrate gains in student achievement. These include the use of growth models to measure progress and the ability to focus intervention and resources on students who have yet to reach grade level (U.S. Dept. of Ed., 2007, p. 7).

As for students with disabilities the DOE further proposed:

To ensure districts receive credit for their work in helping these students make academic progress, states will have the option of assessing a small group of
students with disabilities based on alternative and modified achievement standards. These standards must meet high quality standards and promote challenging instruction so students with disabilities can reach the highest possible levels of achievement (U.S. Dept. of Ed., 2007, p. 8).

It is important to note that DOE has in fact been more lenient since NCLB’s enactment with regards to helping states reach their AYP predictions, but the motive behind this is questionable. Further debate and research is needed to determine if lowering the standards and making it seemingly easier for states to meet AYP predictions is just a move to appease states from ensuing legal actions or if the federal government realized that perhaps its proposed standards were unreachable when first developed. Although the motive is questionable, DOE truly has been more lenient in assisting states and their resident districts and schools in meeting AYP. This must be heavily considered as a positive development for the nation.

There is a continued long list of leniency measures to examine. In 2004, DOE gave “a number of states approval for the practice of ‘banking’ test results when students have two or more opportunities to take a test that yields score used in AYP analyses” (Forte Fast & Erpenbach, 2004, p. 4, also discussed in CEP, 2007). This is seen in the case of schools and districts that previously employed high school exit exams for their students and have now transformed these tests to meet NCLB requirements. With many of these previous exit exams, students were allowed multiple opportunities to pass the test in order to graduate high school. Many states and schools have elected to keep this option available to their students for these types of tests. This is extremely beneficial
because it allows students in schools where they are repeatedly taking standardized tests to avoid an unnecessary repeat of a test they have already passed. Having to repeatedly take an unnecessary test results in lost instructional time, which defeats the ultimate goal of NCLB—raising standards and academic achievement for all.

Under NCLB, schools must meet their AYP predictions, but school districts must also meet their own AYP predictions each year. In 2004 DOE changed the requirements for a district to be designated as in need of improvement by now declaring that the district would have to miss its AYP prediction at each school level (elementary, middle, and high school) for two years in a row. States and districts are also allowed to separate mathematics and reading scores for each school (CEP, 2007; Forte Fast & Erpenbach, 2004). In this case, one possible scenario for a district to be determined as needing improvement would be if within that district AYP predictions were not met in reading at each level. This was a significant amendment for states such as Tennessee, in which 92% of the school districts did not meet their AYP predictions in 2003, but under the approved amendment the percent of school districts not meeting AYP dropped to only 65% (Olson, 2004 in Forte Fast & Erpenbach, 2004).

The Building on Results (2007) report states that “growth models allow states to measure individual students’ progress over time, giving schools credit for improvement from year to year and providing another way to show whether achievement gaps are closing” (U.S. Dept. of Ed., p. 7). The use of growth models is particularly beneficial to subgroups’ scores such as those students with disabilities, or ELL students. In the report, Spellings (2007) goes on to note that “a growth model is a tool to achieve proficiency by
2014, not a loophole to avoid it” (U.S. Dept. of Ed., p. 7). This measure of growth models will allow schools with large subgroup populations who have a history of struggling to show improvement in a new way by demonstrating that academic gains are being made in the classroom, even if they are not pronounced by a dramatic rise in test scores. Other self-described leniency measures taken by the DOE include more flexibility for schools by allowing them to choose how to allocate their funds for students in need of improvement as long as the “all-students” group is meeting the proficient target (U.S. Dept. of Ed., 2007). This allows school and districts to distribute more of their funds to the students who are truly in need of extra support and resources for learning, but they must be careful that the instructional levels of the overall school population do not dwindle in any way. This flexibility is designed to help not only schools under school improvement mandates, but also those that have been placed under corrective action by the DOE.

In its *Building on Results* (2007) report, the DOE continues to try positively to improve broad educational goals, such as making a call for improved graduation rates and to ensure that students who are graduating from America’s high schools have been exposed to challenging coursework and are ready to be productive members of the workforce (U.S. Dept. of Ed., 2007). The report also makes a call for rewarding exceptional teaching as well as improving mathematics achievement, particularly in middle schools where the data analyzed by the DOE shows that results are particularly lacking in this area. The report suggests that the curricula being used for mathematics in the middle school area are not producing results and that new measures to be directed by
the DOE need to be taken. There is also an increased focus on science achievement in which science assessments will be given to three grade levels and officially counted in state accountability reports for the 2008-2009 school year. In the report there is a call to improve reading with the Reading First program, which is targeted to all students, especially limited-English proficient (LEP), African Americans, Hispanics, economically disadvantaged, and students with disabilities (U.S. Dept. of Ed., 2007). Finally, the DOE wants to make an increased effort to redouble opportunities for low-income students using measures such as Promise Scholarships that allow students to attend out-of-district public schools or private schools. New restructuring options will be provided for schools marked as needing improvement. However, they are designed to make deeper gains in improvement. Schools marked as needing improvement on the basis of only one failing subgroup will have different options available to them for their restructuring plan. Additional ways to recruit better teachers will become available to schools that are undergoing restructuring operations, and there will be increased financial support for charter schools (U.S. Dept. of Ed., 2007).

Forte Fast and Erpenbach (2004) as well as CEP (2007) have closely tracked the numerous changes to NCLB, particularly its AYP provisions since the inception of the law. Both have noted that the DOE has been vague at times in its responses to states that have proposed plans for amendments to their state accountability plans, and that often DOE gives not respond in, written or oral form, to a requested amendment (CEP, 2007, Forte Fast & Erpenbach 2004). It is surprising that there has not been more publicized concern over the DOE’s lack of response because it appears to be an injustice to states
looking to modify their accountability plans. It is not reasonable to expect states to revise and edit their accountability plans as the requirements of NCLB change and for the DOE to ignore their concerns in such an outright manner.

In his policy brief for the Center for Research on Evaluation, Standards, and Student Testing (CRESST), Linn (UCLA, 2005) commends the effort of NCLB to improve achievement for low-performing students, and reports that some narrowing of achievement gaps has occurred. However, Linn (UCLA, 2005) notes that there are still some fundamental problems with NCLB, particularly with regards to the AYP measures. He argues that although it is helpful to disaggregate data to ensure that disadvantaged groups are making gains and not being left behind, this same method increased the number of ways that a school can fail to meet its AYP predictions for a given year (Linn, 2005). He also adamantly argues that the Safe Harbor provision that assists in making AYP should allow schools to meet only a three percent reduction in the number of students in the below proficient range instead of the current 10% reduction, noting that this measure would “solve many of the problems caused by expanded subgroup reporting, yet still promote significant achievement increases” (UCLA, 2005, p. 7). The notion of Safe Harbor is similar to that of growth models but it only applies to subgroup populations. If a school demonstrates that there has been an increase in achievement (in this case a 10% increase in scores) within a subgroup, then the subgroup can meet AYP. An interesting relationship that Goertz (2005) assessed was that most of the accountability systems used by states do not include subgroup performance ratings. This can be viewed as either an effort to help states meet state requirements more easily or as a
method by which states do not hold their schools accountable for closing achievement gaps. Lee (2004) who has also done significant work analyzing AYP methods as well as Safe Harbor suggests the use of effect size measures to determine practical significance in gains. However, he notes that “while using an effect size metric with scale scores may help set more realistic performance targets and better recognize schools’ academic progress, it is not permissible under the current law” (Lee, 2004, para. 30). The problem with Safe Harbor is that within some subgroups, such as the ELL or LEP subgroups is that students move in and out of the group. Once a child is proficient in English, he or she moves out of the group and becomes part of the general population of students, but a new ELL or LEP student will take their place within the subgroup which makes it difficult to show improvement with such mobility occurring within the group. To combat this problem, the DOE is now allowing schools to count ELL or LEP students who have become proficient in English to remain in the subgroup for AYP improvement measures for two years after they have become proficient in English (Goertz, 2005).

**Is 100% Proficient Achievable?**

By 2014 all students in testing grades must be at least proficient in reading and mathematics. Students must be able to read and do mathematics at or above their grade level. Little has been said about what happens when this measure is perhaps not met, but plenty has been said regarding the concept that 100% proficient is an unrealistic educational goal (Smith, 2005). Smith (2005) determines that the 100% proficient mark is “unworkable” (p. 512), citing concerns for the unevenly distributed disadvantaged students as well as children who live in large cities—where poverty is more likely to
strike—with a slew of other social concerns that affect educational quality. She concludes that this probable lack of attainment of 100% proficiency will show that NCLB will further enhance the inequities seen throughout American culture and unfairly label children and schools as underachieving (Smith, 2005, p. 513). Linn (UCLA, 2005) notes that “the most serious problem is that NCLB expectations for student achievement have been set unrealistically high” (p. 2; Hill & Barth, 2004) and continues in his CRESST policy brief that he predicts that almost all schools will fail to meet AYP and NCLB requirements within the next few years, long before the 2014 deadline. He makes this claim by employing National Assessment of Educational Progress (NAEP) test results and noting that based on current scores and trends the marks on the NAEP results are far behind the optimistic mark that NCLB sets (UCLA, 2005). He determines that the gains seen on NAEP results should be the same as those reflected on state tests, but based on his analysis on this data, the NAEP test scores are not rising at the same rate as are the state test scores (UCLA, 2005). If one analyzes the available data from the National Center for Education Statistics (NCES) using the NAEP Data Explorer (see http://nces.ed.gov/nationsreportcard to view data in full extent) to examine the NAEP average test results from the 1970s to the 2004 results, one will see that over the 26 years that the math scores have been tracked, 4th graders have increased from an average of 219 points out of a possible 500 to 241 points between the years of 1978 and 2004. The 8th graders have made less improvement, increasing the average score by only 17 points from 264 points in 1978 to 281 points in 2004. The long-term NAEP data provided for
these scores shown in Table 3.1 show that the average either remained the same or increased over the years. There was never a decrease.

<table>
<thead>
<tr>
<th>Year</th>
<th>4th Grade</th>
<th>8th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td>219</td>
<td>264</td>
</tr>
<tr>
<td>1982</td>
<td>219</td>
<td>269</td>
</tr>
<tr>
<td>1986</td>
<td>222</td>
<td>269</td>
</tr>
<tr>
<td>1990</td>
<td>230</td>
<td>270</td>
</tr>
<tr>
<td>1992</td>
<td>230</td>
<td>273</td>
</tr>
<tr>
<td>1994</td>
<td>231</td>
<td>274</td>
</tr>
<tr>
<td>1996</td>
<td>231</td>
<td>274</td>
</tr>
<tr>
<td>1999</td>
<td>232</td>
<td>276</td>
</tr>
<tr>
<td>2004</td>
<td>241</td>
<td>281</td>
</tr>
</tbody>
</table>

Note. NAEP scores are based on a 500 point scale.

Table 3.1: NAEP National Average Test Scores-Mathematics.

The NAEP reading scores paint a very different picture. What is startling about these scores is that America's 4th grade students have never, on average, scored better than 50% correct on this test. The 8th grade students have always achieved just over 50% correct on this test, but it helps the critics of "100% proficient" prove their point that achieving this remarkable percentage in 12 years seems nearly impossible if over a 26-year period scores on a nationally used exam have improved only minutely and have hovered around the 50% mark.
<table>
<thead>
<tr>
<th>Year</th>
<th>4th Grade</th>
<th>8th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>208</td>
<td>255</td>
</tr>
<tr>
<td>1975</td>
<td>210</td>
<td>256</td>
</tr>
<tr>
<td>1980</td>
<td>215</td>
<td>258</td>
</tr>
<tr>
<td>1984</td>
<td>211</td>
<td>257</td>
</tr>
<tr>
<td>1988</td>
<td>212</td>
<td>257</td>
</tr>
<tr>
<td>1990</td>
<td>209</td>
<td>257</td>
</tr>
<tr>
<td>1992</td>
<td>211</td>
<td>260</td>
</tr>
<tr>
<td>1994</td>
<td>211</td>
<td>258</td>
</tr>
<tr>
<td>1996</td>
<td>212</td>
<td>258</td>
</tr>
<tr>
<td>1999</td>
<td>212</td>
<td>259</td>
</tr>
<tr>
<td>2004</td>
<td>219</td>
<td>259</td>
</tr>
</tbody>
</table>

Note. NAEP scores are based on a 500 point scale.

Table 3.2: NAEP National Average Test Scores-Reading.

The long-term trend data has been collected on the NAEP reading test since 1971 and is depicted in Table 3.2. During the 33 years between 1971 and 2004, the 4th grade students only improved 11 points with various increases and decreases in scores throughout the years studied. More disturbing in this data is that the scores are so low. In 1971, the average score on the NAEP Reading assessment (also out of 500 points) was 208 (a mere 41.6% correct), and in 2004 the average score was 219 (only 43.8% correct). In 33 years, the national average scores on the reading portion only improved an overall 2.2%. The 8th graders provide a much more distressing scenario. In 1971 the average reading score on the NAEP was 255. In 2004 the average score increased only four points—to 259. The same fluctuation seen in the 4th grade reading scores is also apparent in the 8th grade reading scores as well. The significantly higher overall increase in math scores could be explained by the country’s emphasis on teaching math and science with more laborious effort on the part of educators, but the overall gain in math scores,
regardless of grade, was still less than one point annually. The progress desired by the federal government can not be made with such mediocre gains. Although the NAEP data is not used to make AYP decisions, it is a valuable tool in examining the progress of NCLB as an educational reform. The NAEP test, although broad in content and level of difficulty, is designed to be a compilation of measures and standards used throughout the United States. Analyzing these national average scores is valuable to defend why an extensive and rigorous reform program such as NCLB is needed, but it might prove beneficial to those wanting to prove that a “100% proficient” mark is an unreasonable goal in such a short amount of time.

Moores (2005) also calls attention to a critical flaw of the NCLB law being the unattainable goal of 100% success for America’s 50 million children and draws special attention to the students with disabilities category. Moores (2005) contends that unlike other special populations mentioned within the actual legislation of the NCLBA, such as the Alaskan Natives, the special education population has no provisions that address it specifically. This indeed important, significantly educationally challenged population that the government has most definitely defined requires some special treatment or modifications as seen in the passing and reauthorization of IDEA. This population has been lumped together with the overall populations of students within the act. The special education community is recognized in particular parts of the Act, such as that dealing with AYP, and has received special attention for their needs, but provisions to assist this group or to make other accommodations for them are few and far between. Finally, Jehlen (2003) notes that the California Department of Education projected that by 2014—
the federal deadline for 100% proficiency for all students—that 98% of its schools will have failed to reach AYP at some point. This self-projection of failure has resulted in many states lowering their standards of “proficient” in order to make a realistic attempt at meeting the total proficiency level by 2014. This effort by states to lower their standards should be questioned more closely by the public since the goal of NCLB is to raise standards. For a state to determine that it will not meet the 100% proficient mark by 2014 should not translate to lowering achievement goals but rather, as the Act intends, the goal should be to raise teaching standards and improve teaching practices to align with the multiple ways that students learn. Peterson and Hess (2005), who graded schools based on the strength of their state proficiency standards, noted that only some states have risen to the challenge and implemented demanding proficiency standards. Many others have lowered standards to inflate scores and make their performance look better. This notion is similar to the Lake Wobegon effect. In 1987 John Cannell published a study that stated that each state in the nation reported their test scores to be above average. Cannell noted that this idea is statistically impossible, and that naturally there will always be some people above average and some below average, but not everyone can be above average. One of the main purposes of the Act is to make schools accountable for their students and what they know, and when shown to have a deficiency in some area, schools are to improve their efforts and states are to oversee this reform, not to find loopholes for schools to continue their failing practices by lowering the standards they must meet. The DOE is heavily criticized for making changes to AYP regulations and perhaps making it easier for states to make AYP, but this effort of states to lower
their proficiency standards is just as appalling because combined with the actions of the DOE it is a double strike against the educational goals that the country has committed itself to.

_Sanctions Imposed on Failing Schools_

As noted previously, when a school or district fails to meet AYP by any of the various measures in any of the various subgroups for two consecutive years the school or district, if designated as a Title I entity, is placed under sanctions and described as in need of improvement. If the school or district is not part of the Title I group, it is issued a warning. At this beginning stage, schools have options for providing supplementary educational services such as tutoring or other programs, and students from families with a low SES are afforded the opportunity and first-choice preference to attend a better-performing school. After five years on the needing improvement list with no measurable improvement in test scores, the school or district begins corrective action, at which time the state can choose to take over the school and replace staff if deemed necessary. The DOE has attempted to be more lenient in this area by first providing schools and districts many ways to make AYP, as seen in all of the changes that have occurred in the requirements for meeting AYP. Providing supplemental services for students costs a considerable amount of extra money, for which schools and districts have fought for, and it was not until recently that the DOE decided to attempt to aid schools more equitably.

In the _Building on Results_ (2007) report, Spellings notes that there will be a “substantial increase in funds for Title I high school students” (p. 10), which aid the students of families who are marked as economically disadvantaged. Currently most Title I funds go
to the elementary schools, not the high schools. To ensure that funds are not taken away from the elementary schools that heavily depend on them, “districts will have to give their high schools at least 90% of the high schools’ proportionate share of the increased funds” (U.S. Dept. of Ed., 2007, p. 10). This financial effort should provide better learning opportunities for students and professional development opportunities for teachers and staff at these schools. By pouring more money into the Title I high schools that will suffer from imposed sanctions if they fail to make AYP, unlike the non-Title I high schools which only receive a warning, more and more of these high schools should theoretically start to make AYP with more ease and not incur the added costs that come with imposed sanctions.

Many people are unaware that prior to the enactment of NCLB, 35 states were already imposing standards on schools whose students were failing to achieve prescribed proficiency standards (Mather, 1999 in Buchman et al., 2000, p. 1-2). And many of the sanctions imposed by these states are mimicked by NCLB, such as the development of an improvement plan, placement on a list of low-performing schools, the option to transfer to a better school, and reorganization (Buchman et al., 2000). Keller and Blair (2003) note that by marking schools as in need of improvement, and therefore driving some students away, the closing of achievement gaps might become harder to obtain because these low performing schools will be filled with students who have no other options and are struggling. Just as seen in the unfunded mandates debate, the federal government argues that many elements of NCLB were already in place in states at the discretion of their own will and power, so “added costs” of sanctions should not be new because so
many states were already using them, but perhaps not as widely throughout their school districts.

The DOE's effort to become more lenient not only in AYP regulations but also in the sanctions imposed on failing schools can be seen in two vastly different lights. One can view the DOE's efforts as humanitarian, and realizing the scope of the problem, particularly with money, and attempting to reconcile the difference between some of the expected extra costs of NCLB and the unforeseen extra costs of NCLB, such as those that arrive when sanctions are imposed, and even with the provided federal funds a school or district cannot fully meet the requirements under their specific sanction. The other way to view this softened leniency of the DOE is that over the past five years, despite the improvement in test scores that the DOE insists on, it realizes more and more that schools will start to be put under sanctions as the expected gains are not met. The DOE's effort to make sanctions less restrictive might really be an effort at damage control. If NCLB does fail in 2014, and 100% of America's students are not proficient in reading and mathematics, or even anywhere near 100% proficient, the DOE will come under harsh attack from the public and it will need to provide some evidence in its defense that it attempted to help schools, districts, and states by every means possible.
CHAPTER 4

STATE STANDARDS

Under NCLB, states must meet guidelines for their state standards for education in each subject area set forth by the DOE. This measure allows for a very broad comparison of all states' standards and as Heise (2006) notes, it might have been a "necessary political price for NCLB's passage" (p. 143). The guidelines set by DOE are indeed very general, which still allows for states to maintain the desired local control over education, which has become a source of fierce debate because it is unclear how comparisons between states can be made when there is so much room for discrepancies between the differing standards of each state. The NCLB legislation states that states must adopt "challenging academic content standards" (NCLBA, 2002, § 1111(b)(1)(A)) that "specify what children are expected to know and be able to do; contain coherent and rigorous content; and encourage the teaching of advanced skills" (NCLBA, 2002, § 1111(b)(1)(D)(i)) followed by the implementation of challenging assessments to test student knowledge. There are no definitions to determine what qualifies as "rigorous content" or "advanced skills." "Although state standards must be 'challenging,' NCLB essentially leaves it to the states to establish their own standards and assessments, as well
as proficiency thresholds” (Heise, 2006, p. 135). The lack of information provided by the legislation is troublesome because it hinders meaningful comparisons between states that could yield information on what states are achieving more success in certain areas, such as with LEP students or members of the special education community. NCLB requires that states undergo a peer review for their accountability plans, which is done via states providing their content standards. The CEP (2007) report states that this single measure helps to demonstrate that state testing systems have become more similar, yet also points out that these systems still vary greatly in the material tested, difficulty of the tests, rigor of the content standards, proficiency thresholds, and ambitiousness of AYP targets. This measure enacted by the DOE under NCLB makes no attempt at leveling the academic playing field for America’s schools. Furthermore the DOE is empowered by NCLB to deny state accountability plans if such elements such as the curricula being used in schools is not based on scientific research. McDermott and Jensen (2005) refer to a New York City public school district that had to throw out an entire curriculum restructuring plan because it was not scientifically valid by federal standards. Some view that “the scientific basis requirement, however disguised as a merely pragmatic, technocratic point, marks a major extension of federal control over curricula and a concomitant reduction of local educators’ autonomy” (McDermott & Jensen, 2005, p. 47). States around the country might show improvement in their test scores, and by 2014 each state might very well reach 100% proficiency, but in theory under these guidelines, we could have 50 different states with 50 different academic standards and 50 different proficiency levels, and we will have done nothing to improve America’s educational status to a more unified
standard, or be able to determine if these students are at the same 100% proficiency level. No valid statement about America’s youth being 100% proficient in reading and mathematics can be taken seriously if we cannot articulate what that really means in terms if academic abilities and strengths. Under NCLB, Congress has allowed the DOE to tread a very fine line of authoritarianism by not making stipulations on states, districts, and schools too harsh, but still leaving remarkable room for differences among states, which further trickle down to the districts and their schools. With regards to each state’s tests “NCLB requires that its test align with state standards so that it will be apparent whether children are truly learning and making real improvement where performance has been inadequate” (O’Neill, 2003, p. 633) and “tests must be aligned to challenging state content standards and meet a wide variety of other criteria” (CEP, 2007, p. 2). There is tremendous importance for tests used within states to be adequately aligned to state standards, and from these tests, assuming that they are valid and reliable, states can measure how their schools are performing based on the set standards, but that same generalization cannot be made against other states that have varying standards that could be significantly different. This is not to say that a national curriculum is needed that determines specifics of what each child needs to know, but in the interests of making meaningful and appropriate achievement comparisons there needs to be more specific generalizations for states to follow. There are organizations such as the National Council of Teachers of Mathematics (NCTM) that work to provide the Principles and Standards for School Mathematics, which is updated periodically that states can align themselves with; this does provide some remedy to the problem because the standards produced by
NCTM are highly regarded and often adopted by states. However, standards provided by the NCTM or any other valid education agency or organization are not considered the law and these do not need to be followed. Hence the issue for legitimate comparisons of state academic standards or even assessments is still very much warranted. A strikingly important result of the lack of alignment between state standards due to the lack of federal regulation is that states cannot share tests, putting an unprecedented burden on the testing community (Harris, 2006). The CEP (2007) report also notes that “most states have undergone a peer review, required by the DOE to ensure that their tests are aligned with standards and meet other technical requirements” (p. 3). These state tests, although being used for the same purpose of accountability, still vary greatly on what material is tested, the content standards being used, as well as cutoffs for proficiency levels in each grade and each subject (CEP, 2007). Linn (UCLA, 2005) also makes note that the proficiency levels on state assessments vary considerably so an accurate comparison is difficult.

Moores (2005) notes that the second major flaw of NCLB is that it allows states to develop their own testing program so there is no reliable mechanism to compare scores, and therefore students and educational achievement among states. Moores (2005) also points out that because every state is left to its own devices, it is the equivalent of having “over 50 NCLBs with common features” (p. 78). This is perhaps why the DOE has mandated NAEP testing of all 4th and 8th graders to allow for some comparison. The link between state standards and state assessments is critical because as Heise (2006) states “the prospect of adverse consequences to states and local school districts flowing
from NCLB induced some states to roll back their student standards” and that even further instead of NCLB creating a “race to the top” it will create a “race to the bottom” (p. 144).

So What About the NAEP Test?

In an attempt to discourage the criticisms of educators about a system of unequal comparisons, the DOE began using the NAEP test many years ago when the trend in educational reform was based on improving test scores. Heise (2006) notes that the NAEP test has been described as too rigorous and too broad, also noting a trend that students tend to do worse on the NAEP than on their state test. The issue of the NAEP test teeters on the line of a national curriculum, which is perhaps only a small but significant step away in our future after the total outcomes of NCLB are measured. The DOE has imposed the use of NAEP in order to make state comparisons, but it is important to note that under the NCLBA law there is no federally required test (Heise, 2006; McDermott, 2003; O’Neill, 2003). This is actually one way in which the federal government dodges legal challenges. It allows NCLB to be firmly followed and regulated by the DOE, but that it does not impose any outrageous testing measures or academic standards. Heise (2006) admits that NAEP does allow for some sort of comparison among states, but it really only helps to prove that there might indeed be significant differences among state standards, tests, and proficiency levels. After viewing the 2005 NAEP data for the 4th grade reading assessment (available at http://nces.ed.gov/nationsreportcard/states), the national average on the test was only 217 out of 500 points. The lowest state average among the 50 states and the District of
Columbia (D.C.) was actually from D.C. with an average of 191. This means that on average, 4th grade students in D.C. answered only 38.2% of the questions correctly on the exam. The highest state average was achieved by Massachusetts, which had an average of 231 points. This is still an appalling score when one considers that it means that on average students in Massachusetts answered only 46.2% of the questions correctly. In Linn’s (UCLA, 2005) CRESST policy analysis he compares proficiency scores on the NAEP test, and when making any inferences on proficiency scores knowing that “proficient” must lie somewhere around the national average, it is inexcusable to think that America accepts “proficient” scores in the 50th percentile, allowing students to essentially know only half of a given subject. In the CRESST policy brief, Linn (UCLA, 2005) notes that the NAEP score results display a narrowing of some achievement gaps that have been present for many years. If the DOE’s main goal is to show that achievement gaps are narrowing, and that scores are rising in general than the NAEP can be used accurately, but most educators are more likely to be interested in comparisons among states to determine who is lacking, just as NCLB is used to compare schools and districts within a single state. The Building on Results (2007) report reaffirms a call to continue the testing of grades three though eight and once in high school to measure achievement and to use the NAEP in grades four and eight to measure national progress, and to continue to disaggregate results to ensure the disadvantaged students are making the necessary academic gains.

Phillips (1993) presents a thoughtful and detailed discussion on using national comparison tests saying that:
A major problem with this method of obtaining national norms is that the statewide and national tests typically differ significantly in content and difficulty. While the statewide test focuses on state-specific instructional objectives, the national test focuses on a much broader content domain common to the textbooks and curricula from multiple states and grade levels. Thus, the coverage of the national tests is typically much broader than that of the statewide test, and the national test contains fewer items per objective. In addition, the national test usually has a broader range of item difficulties and a greater number of more difficult items than the statewide test. Together, these differences decrease the accuracy of the scaling, because the statewide and national tests are not parallel forms and are testing somewhat different levels of difficulty (p. 24).

Solutions to this issue of national comparisons are few and perhaps there is no agreeable answer now. Heise (2006) notes that “indeed, the absence of a common testing metric and proficiency standard continues to frustrate comparisons between or among states” (p. 6), but until a solution is devised that pleases the masses as well as maintains some academic credibility, the use of the NAEP tests as a national comparison test seems to be here to stay.

A very recent development for state-by-state comparisons was reported in the Building on Results (2007) report in which the DOE released plans for a new system for comparing scores based on the tests used by individual states. In order to “promote high state academic standards” states will have to “report the proficiency rates for state and NAEP assessments on the same public report card. Further, the Department of Education
will support cross-state comparisons by providing a platform for states and the general public to analyze and compare standards across the nation” (U.S. Dept. of Ed., 2007, p. 6). Which, specific plans for this program were not reported, it is implied that the federal government will undertake the cost of such a program. However, the CEP (2007) report comments that the scores of changes to states’ testing programs over the past five years even makes comparisons within states more difficult let alone between states. This affects the trust that the public can hold about statements made by policymakers as well as educators about how well NCLB is working or not working and whether America’s children are truly making the promised academic improvements.
CHAPTER 5

CAN ONE TEST REALLY PROVIDE SO MUCH INFORMATION?

_Uses of the Tests Generated by NCLB_

The tests that states are implementing in order to meet the regulations of NCLB are used as various determinants of factors. The tests that America's youth are taking are used to make judgments about those individual children and whether they need remedial help or other services. The tests are also used to make inferences about teacher performance, school performance, district performance, and, finally, state performance. In addition to these obvious uses, these tests implemented under NCLB are used to make decisions about historically disadvantaged groups and the academic progress that they in particular are or are not making. McDermott (2003) notes Airasian (1987) who determined that state legislatures often view student assessment results as a means by which the performance of a given system or policy can be monitored. McDermott (2003) strongly believes that this is an efficient way to determine if funding is supporting the goal of increased student achievement. Although the test scores are followed annually to help demonstrate improvement, particularly in the case of the disadvantaged groups, this one single test each year is being used to provide a lot of data that is driving the decisions.
that policymakers and government officials are making about the amount of federal funding to be designated to particular states. It is true that funding is based on compliance with many stipulations of NCLB, but the main goal, and perhaps the most important goal for schools and districts, is meeting AYP each year with test scores. When this goal is not met, it is almost certain that sanctions and the possible loss of funding will occur.

Goertz and Duffy (2003) cite McDonnell (1994) who asserted that:

The growing reliance on a single test raises a second set of issues: whether one test can serve multiple purposes. Policymakers expect one assessment system to provide indicators of the performance of the education system, hold schools and educators accountable for their performance, certify student performance as students move from grade to grade or out of the K-12 education system, motivate students to perform better, and teachers to change their instructional content and strategies, and aid in instruction decisions about individual students (p. 9).

It seems certain that the government has taken the educational theory behind the exit exams of the old days (which were used to determine if students graduating high schools had learned everything that was intended for them to learn) and has applied it to a much larger scope of students, teachers, administrators, and schools. Miksch (2003) cites an August 2002 CEP report which stated that “policymakers should be careful about expecting tests to bring about significant improvements in teaching and learning” (p. 53). Buchman et al. (2000) discuss the testing phenomenon that was brewing before NCLB was even enacted noting that:
Many believe that the current testing phenomenon reflects our misguided obsession with measuring the outcomes of our educational system (student performance) without first dedicating the resources, in terms of time, effort and money, for students to have any real chance of achieving the standards we set for them. Others say that setting standards and attaching high-stakes will drive reforms within public education that will ensure that students do achieve academic proficiency (p. 1-10).

It is certain that the government has chosen to use NCLB to drive standards up and to incite schools and districts to achieve higher levels of academic excellence. One issue that is prevalent when testing regimes take place is the concern for the curriculum in the schools. "Researchers and educators have argued that high-stakes testing and accountability systems narrow curricula and limit teacher flexibility and creativity" (Goertz & Duffy, 2003, p. 8). After being marked as a low-performing school, many schools and their resident districts decided to make curriculum changes to better align the curriculum with the tests in order to help improve AYP scores in the future (Goertz, 2005). This leads to the old discussion of teaching to the test, and whether meaningful learning that will stay with a child for their lifetime is taking place, or if the child is learning skills that will allow him or her to get an adequate or rather, "proficient" score on a mandated test. Lewis (2000) notes that using high-stakes assessments puts "considerable pressure on individuals or institutions to perform well or to raise scores" (p. 4), which helps to defend the notion that these mandated tests put such intensive pressure on schools and districts now that they are more concerned with finding ways to
achieve good scores on tests. Further, Zirkel (2003) cites attorney Alfred Evans, who represented the Professional Standards Committee in a court case involving a teacher who coached his students who were taking the Iowa Test of Basic Skills (ITBS) with testing supplements that students were legally not allowed to see. The attorney, Evans, said “high-stakes testing, with the expectation of notable increases in all students’ scores, applies huge pressures that can have the unintended consequence of causing teachers to, in effect, cheat” (Zirkel, 2003, p. 712). O’Neill (2003) notes that:

Increased pressure to perform does not necessarily translate into improved performance. Indeed, it could be argued that making additional threats for failure to schools that already seems to have demonstrated that they do not know how to adequately stimulate student achievement can be counter-intuitive and may lead to an increase in cheating on all levels (p. 662).

Jehlen (2003) cites a study by scientists at Arizona State University that reviewed data from over 20 years of high-stakes testing and concluded that high-stakes testing does not improve student achievement as there was no trend in the improvement of scores, or even a decrease in scores. The study also alluded to that fact that schools that had scores that improved dramatically also spent enormous amounts of time on teaching to the content and exact word patterns seen on the tests. The authors of the study also believe that high-stakes testing does in fact narrow the content of the curriculum because so much time is spent on test preparation. (For more on the study see www.asu.edu/educ/epsl/EPRU/epru_2002_Research_Writing.htm). The concern for curricular validity arose after the Debra P. v. Turlington case when African-American
students challenged the fairness of the state graduation exam used in Florida, and argued that African-American students did not have the same preparation as their White counterparts (Phillips, 1993). Curricular validity is important for tests, but the tests should not decide the curriculum. Rather, the high state standards should drive the curriculum and the state assessments should be derived from the curriculum. Buchman et al. (2000) note that “legal challenges have been raised when a test has been used for a purpose other than that for which it was designed” (p. 5-5). The tests used for NCLB are theoretically designed to measure student achievement against state standards; however, the government is using these tests results for much broader goals, which include allocation of funds and sanctions for schools not meeting AYP. So it is plausible that if a hidden loophole can be found in the legal language of the NCLB law, a legal challenge might someday develop against this testing regime that is being grossly overused.

Despite the multiple outcomes derived from the NCLB testing regime, Buchman et al. (2000) believe that we must:

Keep in mind the ultimate goal of this effort: to improve student learning. Test-based standards are one measure of student achievement, and they may help to push students, teachers, administrators and school boards toward higher levels of student learning. They will have the most beneficial effect is the process begins with a well-conceived notion of how the test results fit in with other steps to improve student learning (p. 5-10).

Van Patten (2002) stresses that when using high-stakes testing such as that implemented by NCLB, there needs to be close examination of who is designing the tests,
deciding the content of the tests, and how the scores will be determined and utilized. These are important elements because it is unclear in many cases who is deciding what concepts are being tested on the tests and how efficient and capable those with such an enormous responsibility are, particularly with the large influx of states needing new tests so rapidly to meet newly prescribed standards. Specific provisions regarding the assessment instruments used under NCLB are not prescribed so there is considerable room for poor tests to be used. It cannot be denied that NCLB has had some impact on the educational community. “Many believe that rigorous testing policies, such as high-stakes testing, encourage teachers and students to get serious about teaching and learning” (Lewis, 2000, p. 5) and through all of the aforementioned policy changes and legal debates we see the NCLB really has put the pressure on schools to change, but whether for the better or worse is yet to be determined.

Is All of This Test-Use fair?

Lewis (2000) cites Heubert and Hauser (1999) who said:

The important thing about a test is not its validity in general, but its validity when used for a specific purpose. Thus, tests that are valid for influencing classroom practice, “leading” the curriculum, or holding schools accountable are not appropriate for making high-stakes decisions about individual students…(p. 5).

So, although it might be easy and inexpensive to use the tests under NCLB to measure multiple facets of performance, the validity issue expressed by Lewis (2000) and Heubert and Hauser (1999) says differently. The tests used by states as well as the NAEP imposed by the federal government are all high-stakes tests. Phillips (1993) notes that
the general characteristic of a high-stakes tests is the “considerable pressure on individuals or institutions to perform well or to raise scores” (p. 1). We must question the fairness of allowing one test given at some point in the school year to determine so many choices and outcomes for such a vast number of people. Our educational history has taught us that education is best when left in the hands of the local population, and imposition of such strenuous tests by the federal government in no way keeps with the recognized and highly favored trend of the past. Admittedly, there exist many schools and districts throughout the country that might be hurting students’ education, perhaps through no fault of their own, and perhaps federal input is needed to remedy this situation, but through the heavy use of a test that brings such tremendous anxiety among those who take it as well as need to make decisions by it, is in no way equitable. Phillips (1993) notes that the “characteristics of high-stakes assessment suggest that a high level of anxiety is associated with the assessment and its results and that decisions based on the assessment could potentially deprive an individual or institution of something valuable” (p. 1). When schools repeatedly fail to make AYP, the students risk losing educators who care about their interests, and the educators, some of whom have dedicated their entire lives to teaching, are at risk for being replaced by a state official who bases his or her decision on a percentage of proficient test scores. Because the relationship between test scores and funding is so powerful, a lack of performance on tests could be the result of inadequate financial funding that could not provide the necessary educational experience for students. Guilfoyle (2006) believes that “under NCLB, teachers feel great pressure to focus their energies solely on preparing students to excel on standardized tests” (p. 9).
This is an extreme unintended consequence of NCLB. O’Neill (2003) notes that “nationwide there appears to be a growing retrenchment from, and general backlash against, the imposition of high-stakes exams in general” largely because of the “concerns about damage to some students’ self-esteem, and the effects of unrelenting pressure in the classroom” (p. 658). Parents are increasingly realizing that the pressure these tests are putting on the entire educational community is beginning to take away from the educational experience their children should be receiving in a pressure-free environment. O’Neill (2003) raises a concern with the tests implemented by NCLB, stating that:

If it fails, the losses will be significant. Regardless of whether most students benefit from the imposition of high stakes, many students will most likely not receive any benefit, and, consequently, the numbers of students dropping out of school...will continue to rise (p. 662).

Hess (2006) alludes to the fact that critics of high stakes testing also contend that these tests can cause schools to hold low-achieving students back instead of trying to help them reach what should be attainable academic goals. The purpose of NCLB is to increase educational standards and produce more capable citizens, if the use of these imposed tests cause stress to too many students and they in turn begin to drop out, the objective of NCLB will have failed, and we will have defeated only ourselves.

**Measurement Issues**

The measurement community has provided some special insight into the testing regime imposed by the NCLBA. Harris (2006) explains the appeal of testing under Linn’s (2001) guidance because it is inexpensive, externally mandated, easily
implemented, and is a visible tool to show the public. He draws special attention to the
*Standards for Educational and Psychological Testing* and the influence they have made
on the test-publishing community to ensure that technically sound and psychometrically
appropriate tests are being published, noting that “the *Standards* intends to ‘promote the
sound and ethical use of tests and to provide a basis for evaluating the quality of testing
practices’” (Harris, 2006, p. 45). Harris (2006) also observes that the current level of
educational testing is “unprecedented because testing serves as the linchpin of
educational improvement efforts, attaching very high stakes to test results” (p. 43). As
alluded to previously, states were put under enormous pressure to quickly meet the
demands of the NCLB legislation and align with its regulations, including those for the
assessments they administer. This started a ripple effect that also intruded on the test-
publishing community. All aspects of test construction, from developing questions,
making adjustments for special needs students (including LEP students), as well as the
scoring of the new tests and reporting the scores were rushed, which creates a higher
probability of mistakes being made. Small mistakes can add up to significant losses for
those to whom these assessments are very important. Harris (2006) pleads that the
“compressed timeframes and the expectation that new educational assessments can be
developed quickly without any degradation of their technical quality” (p. 43) creates
competing demands for test publishers. Toch (2006) adds that “many statewide tests are
not getting sufficient psychometric scrutiny to ensure that they accurately measure
student and school performance” (p. 55). Resnick (2006) also alludes to the lack of test
alignment between the standardized assessments and state standards as well as the fact
that the tests tend to omit the more challenging problems that more accurately test students’ problem-solving capabilities.

One key issue particularly related to the assessment aspect of NCLB is that of making special accommodations for students with disabilities. “The allowance to modify assessments undermines the assumptions of standardized administration and generally invalidates psychometric work conducted on the existing assessments” (Harris, 2006, p. 44). The aforementioned notion implies that the modified assessments for students with disabilities require a separate set of studies to determine their validity and reliability particularly in reference to the decisions rendered by these assessments, but there are no provisions within the Act to ensure that this work is done. Harris (2006) also discusses the aspect of universal design, in essence translating high performing and valid tests across different languages and cultures, noting that the number of tests being translated is increasingly rapidly with little or no concern about the technical quality of these revised tests and so therefore accurate inferences about the students taking these revised tests cannot be made.

Guilfoyle (2006) cites Popham (2006) who found that the majority of tests used under the guidance of NCLB cannot detect striking instructional improvements even when they occur. This has two important implications. First, this relates back to the issue of the multiple purposes that a single test is being used for. The tests are supposed to be designed to measure academic achievement, not instructional improvement (at least in the direct sense), so the tests are being used improperly. NCLB is supposed to help produce great academic gains for students through an improvement of educational quality
in the schools. The tests are supposed to be aligned to measure if students are learning the predetermined content set forth by their states. Then some inferences can be made about the improvement in instructional quality. The second issue derives from this misuse of tests to measure instructional improvement; the tests are not designed to measure this explicitly and therefore should not be used in such a manner. Guilfoyle (2006) also discusses the “work-arounds” that researchers have identified that states use to cheat the accountability system and to upset the validity of the measurement system as a whole. Some of these “work-arounds” include making easier tests, lowering proficiency cut-off scores, and discouraging low performing students to attend school on test days. All of these measures negatively affect the measurement system and are producing inaccurate results; furthermore these measures are not helping the students that NCLB was designed to benefit the most.

Another issue that sheds light on the lack of adequate preparation for such an extensive testing regime is the differences in proficiency standards that states used prior to NCLB. Smith (2005) found that students often did quite well on an assessment when graded under the state’s accountability proficiency standards, but did rather dismally under the guidelines of NCLB. Some parents believed their children were achieving excellent test scores and fully meeting state requirements, but then found that the opposite was true under NCLB. Smith (2005) concludes that this “begs the question of whose accountability system most accurately reflects progress and achievement in schools” (p. 512), particularly when these tests being used can produce passing scores by one standard and failing scores by another. Unfortunately, good tests can be fairly
expensive to construct and to score. This relationship between cost and quality is also related to states wishing to meet the demands on NCLB, thus resulting in a desire for tests that measure largely low-level skills. As Toch (2006) notes, “simple tasks are easier and cheaper to test” (pp. 55). He also notes that “most testing experts would prefer to measure students’ grasp of more advanced abilities through open-ended or constructed response questions that require students to produce their own answers” (p. 55-56) but these items are considerably more expensive and time consuming to design and score.

**AYP Differences**

Porter, Linn, and Trimble (2005) examined the consequences of using the various AYP models now allowed by the DOE under the regulations of the NCLBA. They noted that states are allowed to select the trajectories they wish to use to meet the 100% proficient standard by 2014, determine their own minimum N size for subgroup reporting, and choose whether or not to use confidence intervals, and if so at either the 95% or 99% level and whether to use a one- or two-tailed interval (Porter, Linn, & Trimble, 2005). They analyzed data from Kentucky from 2003 and 2004. As expected they found that considerably fewer schools would have made AYP in Kentucky without the use of confidence intervals regardless of the AYP trajectory chosen. Their examination of that data also found that the number of Kentucky schools that would pass AYP increases as the minimum N size increases (Porter, Linn, & Trimble, 2005). States choosing to use AYP trajectories where they are only responsible for score increases

---

6 The DOE allows states to use any of the following trajectory approaches: (a) straight-line approach, where the same increase is seen each year, (b) stair-step approach, where there are constant increases in the required years, (c) front-loaded trajectory, where there are larger increases in the earlier steps, or (d) back-loaded approach, where there are larger increases between years in the final years of the plan (Porter, Linn, & Trimble, 2005).
every three years starting after 2005, such as with the stair-step approach seems most reasonable because it allows schools to prevent themselves from failing consecutive years in a row. However, they must still be making adequate gains so that the new AYP level for the next year can be met. Porter, Linn, and Trimble (2005) note that their analysis of AYP trajectories is important as a measurement issue because it examines the best way for states to meet AYP predictions and hence not be identified as failing, and therefore “may prove pivotal in the years to come in terms of a state’s ability to provide required service to all schools identified and schools’ ability to continue to exist” (p. 39). Resnick (2006) also alludes to the consequence of too many schools within one state not meeting AYP predictions in that “state resources, both financial and human, will have little chance of meeting the assistance responsibility” (p. 34). In reference to Porter, Linn, and Trimble’s (2005) work on states choosing a back-loaded trajectory, Resnick (2006) states:

The choice that these states are making gives them time to hone their systems and learn how to teach all children. However, they may be in for a rude shock later, unless, of course, they are in fact gambling on NCLB being past history by then, or on sufficient pressure building up from states to change the trajectory requirements (p. 35).

The measurement issues whether reflected on test design and scoring, or related to the connection to AYP predictions must be carefully studied to determine the impacts that these decisions have on the students they are designed to help. Resnick (2006) calls
for the tests and measurements to be brought up to par if there is any hope for accountability systems to have any significant effect.
CHAPTER 6

SUGGESTIONS FOR FUTURE EDUCATIONAL POLICY INITIATIVES

The debate over what to do about what Guilfoyle (2006) calls the “United States’ most ambitious federal education law” (p. 8) will continue for years to come, but suggestions for improvement and new directions to move in can indeed be made. As concerned citizens we must consider what NCLB has taught us with regards to the depth of educational policy research, state standards, testing use, AYP, students with disabilities or those considered historically disadvantaged, and the political inner workings of policymaking.

Chester (2006) calls for more research on the debate of the impact of federal law and state accountability systems stating that there are three limitations on the ability of research to inform the future of accountability policy. Chester (2006) states that:

What is offered as research is not always objective….Much of what has been published has been challenged as advocacy rather than research. Second, methodologies for measuring the impact of accountability programs are emerging and often not mature….Third, the boundaries between the research and policy communities are at best semipermeable and subject to reinforcement by the respective parties (p. 3).
Finding refereed research on the many aspects of NCLB is difficult, and it is more difficult to find research that applies to a field, such as policy and its many components, that does not have clearly defined laws, and is also always changing. McDermott (2003) suggests that “scholars should pay more attention to the role of state education agencies in policymaking, and the ways in which state agency capacity may affect policy outcomes” (p. 173) and how state education agencies can help lead reform in the future. Federal policymaking has a trickle-down effect that has historically left many important decisions in the hands of state officials. This was also the case with NCLB. States need to recognize and take hold of the power they have and find ways to use it appropriately, but at the same time the federal government needs to maintain a safe distance that allows states to make the best decisions for the education of their students. Mathis (2005) calls for more scholarly research on the issue of NCLB being over-funded and that “adherents of this school of thought consistently contend that education has enough money, but is inefficiently used” (p. 93). The funding debate, although seemingly complete with no successful legal challenges against the federal government, is far from over. One method to help alleviate the conflicting thoughts that circle around this debate would be an extensive analysis of how funding, federal, state, and local alike are being used by schools and to determine if money is being misspent or misappropriated by any party.

NCLB has been useful in painting the chaotic picture that is state standards. We have seen that it is nearly impossible to determine if state standards can align with each other, much less the tests employed within one state. Considerable work needs to be done to bring about some sort of means to compare states’ standards. A national
curriculum might be the answer to the problem with state standards. Heise (2006) contends that an “illusion of local control” (p. 130) has existed for decades, and that states and the federal government really determine the educational content of schools not local school boards, so we need to let go of this façade that we hold on to so tightly that local control still exists to such a significant degree. A specific curriculum is not needed, but rather a broad curriculum that dictates that every high school student must graduate having completed a certain degree of math, science, English etc. Non-specific regulations would allow for states to at least follow the same general curriculum, but still afford them the opportunity to design their own state standards with emphases in the areas they wish. This would also allow LEAs the opportunity to continue to develop and revise their curriculums as they see fit and use the instructional opportunities that they wish. This notion of a national curriculum will also allow for state-by-state comparisons to be made. If every student in American must know calculus then better tests can measure if every student is proficient in calculus and what states might be achieving the most success teaching it to students, even specified subgroups of students. In addition, valuable and meaningful statements can be made about the educational value of a high school diploma and about America’s ability to educate its youth. In reference to reforms for state standards and tests, Guilfoyle (2006) calls for “encouraging innovation so that we can ensure student learning in a way that supports each students’ future success” (p. 11). Policymakers need to use the services of notable educational experts about new ways to determine what students have learned before repeatedly jumping on to the testing bandwagon with each new federal education policy.
“It appears that the capacity of schools and districts to improve instruction and professional learning dramatically has been outpaced by the much faster-moving development of assessments and accountability systems” (Resnick, 2006, p. 37). From this we can hopefully implement future federal education policy with more time for states and localities to adjust to changes within their systems and plan accordingly. Timelines are appropriate and ensure progress, but certain aspects of NCLB, specifically testing, and aligning tests to state standards, were not given the necessary time to develop to ensure quality products. This has also led to somewhat meaningless state standards that are not being measured with any validity or reliability by the state assessments.

The effectiveness of testing under NCLB will take a long time to determine. Testing has been rudimentally employed and many scholars have noted that the tests are not nearly as well psychometrically fitting as they should be. Wanker and Christie (2005) employ the suggestions from the Educational Commission of the States (ECS) that state that federal officials allow for great variation in testing options such as the use of exams that reflect the varying capabilities of students with disabilities or value-added assessments. ECS also recommends that each testing option be documented as valid, reliable, and aligned with state standards (Wanker & Christie, 2005). This would help remedy many of the concerns that the testing and measurement communities have voiced about the neglect that the state assessments have suffered. The CEP will be producing a report this year that will analyze the data from state assessments to determine if the goals of NCLB are being met and if achievement gaps have narrowed at all since 2002. This
report will hopefully add another dimension to the research on NCLB—research that is not disguised as propaganda to gain support for or against the measures of NCLB.

There are various other ideas that we have learned from the testing regime employed by NCLB. In the early stages of high-stakes testing many people wondered about the “perceived ‘hidden agendas’ in statewide tests” (Phillips, 1993, p. 122) and in today’s contemporary society people are worrying about the hidden agendas of NCLB. This public unease with NCLB is not beneficial to its likely longevity, particularly with its reauthorization on the agenda for Congress this year. Prior to NCLB and the presidency of George W. Bush, the Secretary of Education, Richard Riley, argued that states should examine their “testing programs to ensure that high stakes tests do not results in inaccurate measurements, … and a narrowed curriculum” (O’Neill, 2001, para. 103). We need to return to this important notion and it needs to be more thoroughly addressed in the next wave of federal education policy.

We should question whether the tests used in NCLB will someday be legally challenged because their development was so rushed that it is possible that mistakes will be found that will bear significant burdens on schools who might fail to make an AYP mark when in fact they might have adequately met it. With the success of the federal government in the legal challenges related to NCLB, it seems improbable that any challenge against an element of NCLB will ever be successful and as Zirkel (2004) noted, the fix to NCLB will most likely only be through politics.

Finally, Phillips (1993) provides guidance for policymakers of the future, saying “To make a final decision for a particular assessment program, a policymaker must
consider: (1) the specific policy goals of the program, (2) how this decision will affect other decisions that have already been made or will be made in the near future, (3) what the policymaker feels most willing and able to defend in the particular assessment context, (4) the advice obtained from experts, and (5) relevant laws, administrative rulings, and cases in that particular jurisdiction” (pp. 2-3). Phillips made this suggestion for policymakers nearly ten years before the enactment of NCLB, but we see that nearly all of his thoughtful suggestions seem to have been ignored in the development of NCLB.

The AYP debate and the changes that have resulted have probably gone too far at this point and it might be time to take a “hands-off” approach to AYP before confusing the issue any further. However, one suggestion still stands that would not be burdensome or hurtful to existing state plans, Linn (UCLA, 2005) suggests setting more realistic AYP targets that are attainable in the next seven years. As for the future of any educational policy we should have more defined rules for any AYP type of measure and not allow such drastic changes to occur so frequently.

The ECS recommends that the federal government fairly recognize that students with disabilities learn at different rates and that this needs to be accounted for rather than ignored. Also, ECS calls for the federal government to resolve the conflicts between NCLB and the IDEA to better provide a meaningful education for students with disabilities (Wanker & Christie, 2005). Moores (2005) noted the blatant omission of the special education population in much of the NCLBA. Additional educational policy should be careful to include this significant population despite the existence of a separate policy, such as IDEA, for a subgroup defined as needing special protection under the law.
As with any policy, politics play a significant role. Legislation lives and dies because of politicians’ daily decisions. NCLB came under significant political and public scrutiny over the past five years, largely over its fairness, particularly in reference to the funding of the law. Heise (2006) reminds us that NCLB is not a mandate because it is voluntary; states have the option to opt out with the provision that they will lose federal funding for Title I schools. He says that the 100% participation level in a voluntary program begins to erode public confidence in assertions made by educators and politicians who are upset (Heise, 2006). Also, Goertz and Duffy (2003) note that “well-designed assessments and accountability systems can focus attention on schools and students who need the most help, motivate students and educators, and foster the development of better curriculum and instruction. But policymakers must recognize the limits as well as the promise of such policies” (p. 10). And finally, under his excellent analysis of the political federalism of NCLB, Heise (2006) questions whether it is prudent to permit the federal government to impose such critical education policy beyond its financial contribution. Politics and education might not mix as well as most would hope, but together they are here to stay and some formidable coexistence must take shape in the future for policy goals regarding education to be successful.

Lastly, America’s educators, policymakers, and federal officials must all remember that the aim of this federal education policy is to provide students with an adequate education regardless of any historical disadvantage that they might have and regardless of where they live. In order for NCLB to survive, people cannot feel that it makes “impossible or inappropriate demands” (Resnick, 2006, p. 34). If the federal
government continues to wish for public support of such an extreme measure that involves so many hotly contested issues, they must make the stipulations reasonable to the public eye. Additionally, we need better definitions for many terms under NCLB. Proficiency and funding are in need of a single definition that is acceptable to everyone. There also needs to be a more defined system for developing state standards—one that would allow for accurate and meaningful comparisons to be made. One could argue that NCLB should not be reauthorized instead it should be completely discarded. However such an act would be disheartening and unfair to the hard work of educators across America. Instead, we can only learn from the mistakes and successes of NCLB, including those that are yet to be realized. It might be most beneficial to see NCLB to the end and use the important lessons we will have learned to make a better and more comprehensive educational policy for the future, one that will hopefully be to maintain impressive educational standards, instead of seeking to improve dismal ones.
LIST OF REFERENCES


Goertz, Margaret, & Duffy, Mark. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory Into Practice, 42*(1), 4-11.


80


81


