

Factors to Consider When Evaluating School Accountability Results

ALEX DURAN*

ABSTRACT

This paper contains an overview of factors to consider when evaluating the validity and reliability of interpretations and uses of results used for the purpose of complying with the No Child Left Behind (NCLB) Act.¹ A number of factors are identified and used to examine current interpretations and uses of assessment results for purposes of accountability. A concern is that sanctions and consequences may be imposed on schools through the use of invalid and unreliable results. Specific NCLB Act (NCLB) requirements are identified and used to examine this claim. The requirements include: The development and implementation of content and performance standards and standards-based assessments; Adequate Yearly Progress (AYP) with a focus on potential negative impact based on immediate implementation and vague operational definitions; and sanctions and consequences. Clarification of interpretations and uses of results is provided to develop a better understanding for stakeholders who are responsible for making policy and educational decisions. In conclusion, the author suggests that the NCLB Act creates an opportunity for all states to develop and implement valid and reliable accountability systems that clearly and accurately identify effective schools and also provide adequate support to schools in need of improvement so that all students are able to receive quality and effective instruction that improves academic achievement and ultimately allows for students to reach their full potential.

*Alex Duran, Ph.D. is Director of Research, Assessment, and Evaluation at Sunnyside Unified School District in Arizona.

1. Signed by President George W. Bush on January 8, 2002, NCLB provides schools with groundbreaking educational reform opportunities based on stronger accountability for results, more freedom for states and communities, encouraging proven educational methods, and more choices for parents. No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002) (principally classified to Chapter 70 of Title 20, 20 U.S.C.A. § 6301 et seq).

I. INTRODUCTION

The purpose of this paper is to identify factors that are not accounted for but create difficulties with interpretations and uses of results that are generated and used for purposes of school accountability. The importance of the process of validation is repeatedly emphasized with examples of contextual differences that exist between schools to examine the clarity, accuracy, and justifiability of interpretations and uses of results. A primary concern is that some schools face the possibility of confronting consequences and sanctions based on results that may be invalid and unreliable.²

The paper begins with a statement of the problem followed by a historical overview of federal educational funding policies to suggest that the primary purpose for the NCLB Act³ is to ensure that schools are truly accountable for making AYP toward providing all students with quality and effective educational experiences and opportunities that enable them to continually achieve academic success. In addition, NCLB requirements are intended to ensure that historically disadvantaged and special groups of students are also included as part of the vision of improved academic achievement. These thoughts are different from one suggesting that the primary goal of NCLB is to place control of public education in the hands of the federal government.

The historical overview on federal funding is followed by a general overview of NCLB, which is comprised of several requirements but only a select few are used to identify factors that are not included in the methods used for the purpose of determining AYP. The discussion begins by defining content and performance standards to identify their intended purpose. Although the content and performance standards appear to be a logical approach to improving academic achievement, there are concerns toward the likelihood of differentiated school and student impact that will be created by requiring the same standards for all students.

2. Under NCLB's accountability provisions, states must describe how they will close the achievement gap and make sure all students, including those who are disadvantaged, achieve academic proficiency. They must produce annual state and school district report cards that inform parents and communities about state and school progress. Schools that do not make progress must provide supplemental services, such as free tutoring or after-school assistance; take corrective actions; and, if still not making adequate yearly progress after five years, make dramatic changes to the way the school is run.

3. For a comprehensive list of NCLB Act requirements, see <http://www.ed.gov/policy>.

The discussion on standards-based assessments defines the intended purpose for assessments and identifies concerns with the difficulty of making clear and accurate interpretations of results in order to make better use of the information they provide. Difficulties with interpretations are based, in part, on a lack of contextual information to allow one a better understanding of existing differences between schools. Other concerns with high-stakes accountability are the potential for corruption of indicators and the difficulties that educators have in being able to clearly and accurately interpret and use results in meaningful ways. The final topics presented in this section are intended to create an awareness about the power of social and political arenas as well as the power of confirmation bias to suggest that results can be interpreted in a manner that allows for some to drive political policy decisions while allowing others to confirm what they want or choose to believe. Concerns with political policies and biased beliefs are that they are not always accurate or useful.

A brief history on accountability will establish an understanding for evolving changes in AYP, which is proceeded with a discussion on concerns with methods used to determine AYP. It is suggested that perhaps the greatest challenges confronting AYP are unrealistic timelines that were provided for implementation and vague operational definitions. It is also suggested that the method used to determine AYP creates a greater hardship for some schools than for others. A primary concern with the method used to determine AYP is that some schools face the possibility of experiencing consequences and sanctions despite a lack of valid and reliable empirical evidence that clearly and accurately justifies and supports such methods or actions. The final discussion on AYP examines the power of public perceptions and the influence that such perceptions have on public policies and expectations.

The final requirements examined are sanctions and consequences, which were intended to provide incentives for schools to improve instruction and learning. The concept of context is further developed and used to examine the difficulty of being able to reach consensus on interpretations and uses of results in applied context. It is suggested that there are differences in the levels of difficulties related to the variability of school challenges that can better be defined with enriched contextual information. The section ends with a discussion on Opportunity-to-Learn to examine the importance of providing schools with sufficient and quality resources in order to improve academic achievement.

In conclusion, it is suggested that the challenges presented by NCLB are difficult but worthwhile because of the moral and ethical responsibilities that are associated with developing and establishing educational systems where all students receive equitable and excellent learning experiences and opportunities at every school, grade level, and classroom. The journey is not simple as validation is a process that requires continual refinement of ideas, practices, and policies in order to better meet the needs of our most precious resources, our students. If developed and used properly, NCLB can become the vehicle that transforms ordinary classrooms into extraordinary learning environments where all students are able to discover and reach their full potential in order to successfully pursue and achieve their life-long goals and ambitions.

II. STATEMENT OF THE PROBLEM

Today, schools face the possibility of confronting consequences and sanctions based on interpretations and uses of results that may be invalid,⁴ unreliable,⁵ ungeneralizable,⁶ and of little value or utility in providing the type of information that is needed in order to improve curriculum, instruction, and achievement.⁷ These possibilities exist, in part, because NCLB requirements are not based on empirical evidence that

4. R.L. Brennan, *Some Problems, Pitfalls, and Paradoxes in Educational Measurement*, 20 *Educ. Measurement: Issues and Practice* 6, 6-18 (2001); E. H. Haertel, *Validity Arguments for High-Stakes Testing: In Search of the Evidence*, 18 *Educ. Measurement: Issues and Practice* 5, 5-9 (1999); M.T. Kane, *Current Concerns in Validity Theory*, 38 *J. Educ. Measurement* 319 (2001); T.J. Kane & D.O. Staiger, *Volatility of School Test Scores: Implications for Test-Based Accountability Systems. Accountability and Its Consequences for Students: Are Children Hurt or Helped by Standards-Based Reforms?*, Brookings Instn. Symposium on Educ. Policy (2001); R.L. Linn & C. Haug, *Stability of School-Building Accountability Scores and Gains*, 24 *Educ. Evaluation and Policy Analysis*, 29; S. Messick, *Validity*, in R.L. Lynn, *Educ. Measurement* 13, 13-104 (R.L. Linn, 3d ed., Am. Council on Educ./Macmillan 1989). L. Olson, *Study Questions Reliability of Single-Year Test-Score Gains*, *Educ. Week* 1, 1-4 (May 23, 2001); K. Ryan, *Assessment Validation in the Context of High-Stakes Assessment*, 21 *Educ. Measurement: Issues and Practice* 7, 7-15 (2002); R.D. Schwarz et al., *The Challenge and Attainability of Goals for Adequate Yearly Progress*, 20 *Educ. Measurement: Issues and Practice* 26, 26-35 (2001); F. Sicol, *What Do School-Level Scores from Large-Scale Assessments Really Measure?*, 21 *Educ. Measurement: Issues and Practice* 17, 17-26 (2002).

5. T.J. Kane & D.O. Staiger, *supra* n. 4; R.L. Linn & C. Haug, *supra* n. 4, at 29; L. Olson, *supra* n. 4, at 1-4; F. Sicol, *supra* n. 4, at 17-26 (2002).

6. E.H. Haertel, *supra* n. 4, at 5-9; K. Ryan, *supra* n. 4, at 7-15.

7. E.H. Haertel, *supra* n. 4, at 5-9; T.J. Kane & D.O. Staiger, *supra* n. 4; L.A. Shepard, *The Role of Assessment in a Learning Culture*, 29 *Educ. Researcher* 4, 4-14 (2000).

can be used to clarify and justify the accuracy of interpretations for results, which is an important step toward validating their intended use.⁸ The lack of valid and reliable empirical evidence also creates difficulties with the *accuracy* of interpreting results. For example, it is possible for ‘truly’ effective schools to be identified as ineffective while ‘truly’ ineffective schools can be identified as effective. In both cases, the consequences are punitive because the ‘truly’ ineffective school identified as effective is allowed to continue to assume success with implementation of programs intended to improve instruction and learning while the ‘truly’ effective school identified as ineffective can experience unjustifiable sanctions and consequences, which include being forced to modify or eliminate ‘truly’ effective programs. Because it is possible that schools could be misidentified under the current system, it is suggested that empirical evidence be required to clearly and accurately establish the validity and reliability of interpretations and uses for results.

A. Validity

Validation is a process that requires use of empirical evidence to clarify and justify continued or changed interpretations and uses of test results. Positive change is more likely to occur when new or improved empirical evidence is continually gathered, scrutinized, and used for the purpose of improving existing systems. The validation process is difficult because it requires an accurate interpretation of observed scores in order to reach consensus on the appropriate uses of test scores in applied context.⁹ Despite the difficulties, the validation process is essential and must be ongoing in order to evaluate the accuracy and adequacy of inferences and actions that are based on test scores or other modes of assessment in order to continually improve them.¹⁰

Planned or unplanned consequences are also an important part of validity. The unpredictable and unforeseeable nature of intended and unintended consequences requires ongoing evaluations to determine whether, for example, the parental options¹¹ listed under NCLB are

8. M.T. Kane, *supra* n. 4, at 319-342; T.J. Kane & D.O. Staiger, *supra* n. 4; S. Messick, *supra* n. 4, at 13-104.

9. M.T. Kane, *supra* n. 4, at 319-342.

10. E.H. Haertel, *supra* n. 4, 5-9; M.T. Kane, *supra* n. 4, 319-342; K. Ryan, *supra* n. 4, at 7-15.

11. Under NCLB, underperforming schools must use their federal funds to make needed improvements. In the event of a school’s continued poor performance, parents have the option to transfer their child to higher-performing schools or receive supplemental educational services.

empirically based, justifiable, and truly effective at improving academic achievement. A sustained validation process enhances the likelihood of continually refining the clarity and accuracy of interpreted results for the purpose of providing information that truly allows for stakeholders to make decisions that are successfully transformed into improved academic achievement, especially for students who are at-risk of academic failure.

B. Reliability

Reliability is a characteristic of scores that involves quantifying their consistency or inconsistency.¹² Reliability refers to the consistency of scores obtained by the same persons when reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions. The inconsistency of results creates challenges with the clarification and justification of their interpretations and uses. Inconsistencies in school and student results are due to a number of factors that include, but are not limited to, measurement and sampling errors,¹³ non-persistent factors such as distractions during testing, illnesses, personalities, etc. Individual idiosyncrasies can also create fluctuations in the consistency of scores.¹⁴ Fluctuations in school results may also occur because of differences in student characteristics from one year to the next or the total number of students who are tested in a given year. One could expect a decreased reliability of results based on different groups of students. Decreased reliability is also expected when comparing results based on small groups of students. The reliability of assessment results is critical when evaluating the consistency of scores from year-to-year.

III. HISTORICAL OVERVIEW OF FEDERAL EDUCATIONAL FUNDING POLICIES

The primary goal of NCLB has been described as an attempt to place control of public education in the hands of the federal government. This

12. R.L. Brennan, *supra* n. 4, at 295-318.

13. The concept of reliability underlies the computations of the error of measurement of a single score, whereby one can predict the range of fluctuations likely to occur in a single individual's score as a result of irrelevant, chance factors.

14. T.J. Kane & D.O. Staiger, *supra* n. 4.

belief may be attributed to the stringent nature of NCLB accountability requirements.¹⁵ An alternative belief is that the purpose for the NCLB is to establish national policy clearly documenting a genuine national commitment to improving academic achievement for at-risk students, which can only be achieved by accounting for their academic achievement and progress and by establishing a clear timeline for reaching clear goals. The alternative belief is supported by the fact that the federal government has allocated funding for approximately forty years in an attempt to improve academic achievement for disadvantaged students.

In 1965 Congress enacted the Elementary and Secondary Education Act (ESEA) providing federal educational funding primarily for culturally disadvantaged students. In 1982, Congress passed the Education Consolidation and Improvement Act (ECIA), replacing the ESEA with block grants and legislation comprised of chapters. Chapter I—currently known as Title I—was the largest federal fund provided to expand educational opportunities for educationally deprived children.¹⁶ The only eligibility requirement for federal funding was the percentage of students who qualified for free or reduced lunches. Over the course of several years, schools were suddenly required to demonstrate AYP toward improving the percentage of students who were able to demonstrate academic proficiency on standardized assessments. There were no significant consequences for schools that continually failed to make AYP. In addition, English Language Learners (ELLs) and Special Education students were exempt from AYP measures, which meant that there was no meaningful or purposeful accountability for their academic achievement and progress.¹⁷ An important historical fact is that the allocation of federal funding had continually failed to produce intended results for students who were at risk of academic failure and had also failed to account for the academic achievement and progress of ELL and Special Education students.

15. Schools are required to make AYP by improving academic achievement in order to avoid sanctions and consequences. By 2013-2014, 100% of all students are expected to demonstrate academic proficiency.

16. K. Alexander & D.M. Alexander, *Role of the Federal Government*, Am. Pub. Sch. L. 65, 65-66 (5th ed. West Thomson Learning 2001).

17. ELLs and Special Education students were historically exempt from school accountability results. NCLB requirements make schools accountable for all students, including ELLs and Special Education students.

In January 2002, NCLB was enacted to account for improvement in the academic achievement of *all* students. NCLB contains requirements that are intended to have a positive effect on school reform efforts. The requirements are the most stringent ever applied to federal educational reform efforts, in part, because they account for the academic achievement and progress of all students and they include severe consequences for schools that continually fail to make AYP. Although the requirements are stringent, a devil's advocate would suggest that a failure to include more stringent requirements would most likely result in the continued practice of exempting special groups of students from accountability and continued low academic achievement by at-risk students.

IV. NCLB GENERAL OVERVIEW

NCLB contains requirements that are intended to clearly define expectations in order to improve instruction and learning for the benefit of all students. States are required to develop content and performance standards that apply equally to all students. Content standards define the skills and knowledge that all students are expected to obtain and be able to demonstrate while performance standards define proficiency levels for skills and knowledge, which are required to be measured and determined through the use of standards-based assessments. The content standards are required to be taught by highly qualified staff¹⁸ and states must develop and use valid and reliable standards-based assessments to determine how well students in grades 3–8 and high school have learned the required content standards.

Schools are required to make AYP by increasing the percentages of students who are able to demonstrate proficiency on annual standards-based assessments. By 2013-2014, schools are expected to show that 100% of all students are academically proficient. To ensure that all student groups make adequate progress, school results are disaggregated by ethnicity, ELL status, Special Education, and free or reduced lunch—serving as a proxy measure for home economic status. All subgroups must meet the AYP criteria in order for a school to make AYP. Schools failing to make AYP are subject to immediate sanctions and consequences.

18. For more information on qualifications for teachers and paraprofessionals, See <http://www.ed.gov/policy/elsec/leg/esea02/index.html>.

A. NCLB Requirements

1. Content and Performance Standards

Content and performance standards are key components of national accountability and reform efforts. The content standards define the skills and knowledge that all students are expected to know and be able to demonstrate at benchmark grades. All teachers are required to be highly qualified and able to teach the state academic content standards to all students. The performance standards define the levels of students' knowledge (e.g., advanced, proficient, basic), which are to be measured and determined through the use of valid and reliable standards-based assessments. All students are expected to achieve, at a minimum, a proficient level of academic ability. The content and performance standards apply equally to all students as opposed to only a selected subset of college-bound students.¹⁹ Schools are required to make AYP by increasing the percentage of students who are able to demonstrate proficiency from year-to-year. Although having the same content and performance standards is a clear sign of raised expectations in academic achievement, there are concerns about the need to have the same requirements for all students.

2. Concerns with Content and Performance Standards

Requiring the same content and performance standards creates a number of concerns because all students are expected to achieve, at a minimum, a proficient level of academic ability at benchmark grades. This is a difficult task because not all children are able to learn the same curriculum in the same amount of time. The fact is that children in the United States currently do not have equal opportunities to learn or to succeed,²⁰ which creates concerns²¹ about the adoption of accountability systems that require all students within a specific grade level to demonstrate proficiency. Linn suggests the alternative possibility of having high standards without having common standards for all students, especially since they do not possess the same levels of academic abilities

19. M.E. Goertz & M.C. Duffy, *CPRE Policy Brief*, Consortium for Policy Research in Educ. 1, 1-7 (2001); R.L. Linn, *Assessments and Accountability*, 29 Educ. Researcher 4, 4-15 (2001).

20. *Id.*

21. *Id.*

when they begin school. He also suggests that having high standards of performance at a given grade level does not have to be interpreted as having uniform standards for all students, especially since ELLs and Special Education students will most likely progress at a different rate. Other concerns with content and performance standards are in the areas of curriculum and instruction.

NCLB requires that states develop and align content and performance standards with standards-based assessments so that schools can also align their curriculum and instruction with the same standards and assessments that are used to measure achievement in order to determine AYP for schools. A concern with this approach is that standardized tests can also have undesirable effects on teaching and learning because such practices can lead to a narrowing of the curriculum and an overemphasis on basic skills. As a result, the process of alignment must continually be evaluated to ensure that current uses of assessments do not serve a disruptive function for teaching by “teaching to the test,” which translates into teaching a set of skills that have little to do with deep competence.²² The unintended consequence to such an approach is that the teaching of basic skills requiring lower cognitive functions is ineffective as gains typically seen during the first years after a new test is put in place may not transfer to other tests, even those ostensibly measuring the same thing.²³ One final concern with requiring the same content and performance standards for all students is that this approach can also encourage a narrowing of educational experiences for students, greater failure rates, and a failure to develop many worthwhile talents.²⁴

B. Standards-Based Assessments

Historically, norm-referenced assessment results have been used to compare school results to infer different levels of success with instruction and learning. The information has been made public and has been intended to motivate schools with lower scores to improve. This practice persists today despite the fact that a primary function for norm-referenced assessments is to ensure that 50% of scores fall at or above a nor-

22. *Id.*; A.H. Schoenfeld, *Looking Toward the 21st Century: Challenges of Educational Theory and Practice*, 28 *Educ. Researcher* 4, 4-14 (1999).

23. R.L. Linn, *supra* n. 19, at 4-15.

24. W.E. Coffman, *A King Over Egypt, Which Knew Not Joseph*, 12 *Educ. Measurement: Issues and Practice* 5, 5-8 (1993).

mal distribution while the other 50% of scores fall at or below the same normal distribution. This could mean that if the average results for each school were plotted there would most likely be a normal distribution.

Today, standards-based assessments maintain a similar function by serving as the operational arm of school reform efforts, which are intended to direct attention to content and performance standards by including rewards and consequences that serve to motivate student and teacher performance.²⁵ These accountability functions continue to occur despite the fact that a primary purpose for assessments is to provide meaningful and useful information that can be used to improve instruction and learning. The continued use of assessment results for purposes other than what they were intended also creates concerns about the usefulness of variable types of information that are provided. One concern is that the standards movement has been corrupted into a heavy-handed system of rewards and punishments without the capacity building and professional development originally proposed as part of the vision.²⁶ Shepard also emphasizes the importance of recognizing the pervasive negative effects of accountability tests and the extent to which externally imposed testing programs prevent and drive out thoughtful practices. Her concern is that under intense political pressure, test scores are likely to go up without a corresponding improvement in student learning.

1. Standards-Based Assessment and Factors to Consider When Interpreting Results

The process of clearly and accurately interpreting results is difficult. Annual differences in assessment results between schools are commonly interpreted as meaningful differences in their abilities to provide quality and effective instruction, and the same interpretation is also applied to annual changes in a school's results based on year-to-year comparisons of assessment results for different groups of students at benchmark grades. The degree to which a school is looked upon as effective is generally determined by where it ranks in comparison to other schools based on standardized assessment results. One must recognize that the process of accurately determining the level or degree of school effectiveness will most likely not be achieved through rank ordering.

25. E.L. Baker, *Visions of Test Results Dance in Their Heads*, Am. Educ. Res. Assn. Conf., New Orleans, LA (2002).

26. See L.A. Shepard, *supra* n. 7, at 4-14.

School effectiveness is also determined by comparing changes in a school's results from year-to-year, but this method is also questionable because underlying factors remaining unaccounted or controlled for serve to confound the clarity and accuracy of interpretations.²⁷ It is common practice to interpret changes in scores as the positive or negative effects of instruction on learning depending on the direction of change. In addition to change, one must also recognize that it is possible that a school's performance may be unchanged over several years because good instruction occurred but the test was insensitive to measuring change.²⁸

Other factors need to be considered when looking at year-to-year changes in scores to determine school effectiveness, especially when they are based on comparative changes in academic achievement between successive groups of students. Such measures contain a tremendous amount of noise due, in part, to differences between the comparative groups.²⁹ A concern with this method is that such comparisons are most reasonable for schools with consistent student populations.³⁰ The validity of inferences based on year-to-year comparisons for different groups of students is questionable, however, for schools with rapidly changing demographics or with too few students tested in a specific grade. There are alternative methods that can provide better measures for determining school effectiveness and improving academic achievement. Longitudinal measures are one example of such a method as measures of individual student progress have much more potential for educational benefit than measures for different groups of students.³¹

Changes in results are also due to factors that may not be within the control of schools. Such factors include: (1) the particular group of students tested in a given school year,³² (2) measurement and sampling

27. R.L. Linn & C. Haug, *supra* n. 4, at 29-36; J.W. Popham, *Where Large Scale Educational Assessment is Heading and Why it Shouldn't*, 18 *Educ. Measurement: Issues and Practice* 13, 13-17 (1999); K. Ryan, *supra* n. 4, at 7-15; R.D. Schwarz et al., *supra* n. 4, at 26-35; F. Sicolu, *supra* n. 4, at 17-26.

28. E.L. Baker et al., *No Child Left Behind*, CRESST Line 1, 1-8 (Spring 2002). CRESST Line is the Newsletter of the National Center for Research on Evaluation, Standards, and Student Testing.

29. R.L. Linn & C. Haug, *supra* n. 4, at 29-36.

30. R.L. Linn, *Reporting School Quality in Standards-Based Accountability System*, CRESST Policy Brief 3, 1-4.

31. R.L. Brennan, *supra* n. 4, at 6-18.

32. T.J. Kane & D.O. Staiger, *supra* n. 4; R.L. Linn & C. Haug, *supra* n. 4, at 29-36.; F. Sicolu, *supra* n. 4, at 17-26.

error,³³ (3) non-persistent factors,³⁴ (4) school size,³⁵ (5) the corruption of indicators,³⁶ (6) other confounding variables,³⁷ and (7) a lack of contextual information when drawing inferences from test performance to school quality.³⁸ Poverty and nutrition are examples of contextual information that serve to enhance an understanding for differentiated school results.³⁹ Such contextual factors can and do have a profound impact on student and school achievement. Knowing that numerous factors play an integral part in the outcome of results will enhance the processes of validation and sound decision-making.

2. Standards-Based Assessment and the Importance of Context with Interpretations

Variations in the academic achievement of students that occur between schools can be explained by a number of factors that also serve to define differences in the contextual climates and challenges of individual schools. In cases where minority students attend homogeneous schools,⁴⁰ it is highly likely that they will comprise a small minority group attending an Anglo majority school where they are least likely to fit in or be accepted if they are poor or ELLs. Minority students are more likely to attend heterogeneous schools where there is greater variation in the ethnic composition of students and their home economic status. In cases where minority students are the majority group, there are likely to be a large number of students on free and reduced lunch, which serves as a proxy measure for poverty. Certain factors generally have the profound effect of placing some schools at a disadvantage, particularly those serving a significantly large number of poor students because they are less likely to perform well on measures currently used for purposes of accountability.⁴¹ The common occurrence of such outcomes is well

33. R.L. Linn & C. Haug, *supra* n. 4, at 29-36.

34. T.J. Kane & D.O. Staiger, *supra* n. 4; R.L. Linn & C. Haug, *supra* n. 4, at 29-36.

35. *Id.*

36. See R.L. Linn, *supra* n. 19, at 4-15.

37. H.G. Glidden, *Breakthrough Schools: Characteristics of Low-Income Schools that Perform as Though They Were High-Income Schools*, 17 *J. of Sch. Research and Info.* 21, 21-26 (1999); J.W. Popham, *supra* n. 27, at 13-17; K. Ryan, *supra* n. 4, at 7-15.

38. K. Ryan, *supra* n. 4, at 7-15.

39. H.G. Glidden, *supra* n. 37, at 21-26; J.W. Popham, *supra* n. 27, at 13-17.

40. In homogeneous schools, students are more alike than different from one another.

41. F. Sicol, *supra* n. 4, at 17-26.

documented⁴² as racially integrated schools are less likely to win awards for academic achievement when compared to racially homogeneous schools comprised primarily of non-minority students. Unfortunately, current accountability systems do not include methods for adjusting school accountability results based on differences in the ethnic, language, and economic composition of students. On the other hand, one could also suggest that any adjustments would only perpetuate the continued academic achievement gaps that have long persisted between minority and non-minority students.

3. Standards-Based Assessment and the Potential for Corruption of Indicators

The use of assessments in high-stakes environments creates a number of other concerns with interpretations of results as research⁴³ has shown that standardized test results in high-stakes accountability systems has also yielded inflated impressions of student achievement. Such outcomes were attributed to the deliberate corruption of indicators, particularly when tests were used for accountability or other high-stakes purposes. Examples of corruption include: (1) School personnel learning how to code assessment answer sheets in order to select the list of students who are aggregated into the school results for purposes of accountability, (2) over identification of students into special education programs to create exemptions from accountability, and (3) a greater likelihood of classifying ELLs as special education students once their English language exemption windows expired. The examples provided include two specific groups of students who have historically been exempt from accountability, which has also meant that their rate of academic achievement and progress has not been accounted for. Although teaching becomes a greater challenge when external factors serve as the prevailing forces that create delays in academic growth and academic achievement for some students, such factors must not become the readily accepted excuse for justifying lowered expectations, standards, or opportunities for certain groups of students. Accounting for the academic achievement and progress of all students can only enhance the likelihood that special needs will be met in a manner that results in improved academic achievement.

42. T.J. Kane & D.O. Staiger, *supra* n. 4.

43. R.L. Linn, *supra* n. 19, at 4-15.

4. Standards-Based Assessments and the Ability to Interpret and Share Results

Educators play a critical role in defining and determining the accuracy of interpretations and uses of assessment results. There are a number of factors that will determine the degree that educators can and will be able to accurately interpret and effectively use assessment results in order to improve instruction and learning. First, educators must clearly understand the intended purpose for specific assessments and be able to use the information that is provided to improve instruction and learning. Understanding the purpose for different types of assessments also enables one to recognize the strengths and the limitations of specific assessments and the results that they provide in order to clearly and accurately determine when and how best to use the information. Educators must also be able to clearly and accurately interpret and share assessment information with other stakeholders who are also responsible for providing assistance in the process of improving instruction and learning. The process of fulfilling these objectives will be difficult as research⁴⁴ shows that experienced teachers might not understand the benefits of different types of tests (e.g., norm-referenced v. criterion-referenced) or the differences in their applied use. The inability to accurately interpret and effectively use results further inhibits the likelihood that parents and students will learn how to interpret and use assessment results in order to assist teachers with monitoring and improving academic achievement. The impact will most likely be greatest for at-risk students and their parents because they will experience the greatest challenges in learning how to interpret and use this information in order to obtain the greatest benefits.

5. Interpretations of Results and the Power of Social and Political Arenas

Interpretations and uses of assessment results also extend beyond the classroom into social and political arenas. State educational agencies are generally responsible for providing interpretations and uses for assessment and accountability results. A concern is that the interpretations and uses of results are commonly prescribed by the same state educational

44. B.B. Palmer, Paper, *Factors Affecting Educators' Belief Systems Regarding Tests*, Ariz. Educ. Research Org. Conf. (Phoenix, 2002).

agencies that are also responsible for developing and implementing methods for determining AYP, which could create a conflict of interest. In addition, the methods that are developed and used to determine AYP are generally not clearly understood by the general public, thus making it difficult for them to clearly and accurately determine whether or not they truly produce valid and reliable results. How well the general public understands the methods for determining AYP is often dependent on the clarity and accuracy of interpretations that are provided by the state. This creates the possibility that the accuracy of interpretations may vary and can be tailored for the purpose of gaining the public's support in accepting the prescribed and intended uses of results. If interpretations of results are not based on empirical evidence and expert scrutiny, then it is difficult to determine validity and reliability. A lack of empirical evidence and objective professional scrutiny creates the possibility that public information may be tailored in subtle and perhaps not so subtle ways for the purpose of gaining social and political acceptance for prescribed interpretations and uses of results that are described as 'fair' and 'accurate' as opposed to 'valid' and 'reliable.'

6. Interpretations of Results and the Power of Confirmation Bias

Factors such as confirmation bias may explain why clarifications and justifications that are readily provided are easily accepted as clear and accurate interpretations and uses of assessment results. Confirmation bias is defined as the tendency to look for evidence that supports the assessment enterprise.⁴⁵ For example, schools serving students in high socio-economic neighborhoods are expected to obtain higher test scores than schools serving student in lower socio-economic areas. Differences in results between such schools are used to confirm what policy makers and affluent parents want to believe; higher test scores are a clear indication of a better school. Similarly, schools are more likely to attribute positive changes in their results to positive changes in the quality and effectiveness of instruction and are just as likely to attribute negative changes in their results to factors beyond their control. State accountability systems and policy makers are likely to be guilty of confirmation bias because of the intense political power that affluent parents have and the difficulty in being able to accurately determine the total number and

45. K. Ryan, *supra* n. 4, at 7-15.

types of schools and students that the general public will allow to endure severe sanctions and consequences.

C. Accountability and Adequate Yearly Progress

A brief history on federal accountability shows that assessment results have long been used to evaluate programs and schools. The historical timeline also shows how the concept of AYP has evolved. In 1965 the ESEA provided federal funding for Chapter I in order to provide supplemental educational services and resources to students who were at risk of academic failure. Schools receiving federal funding were required to provide assessment results to evaluate their programs.⁴⁶ In 1994, changes were made to the Improving America's Schools Act (IASA), which required states to measure student performance at least once between grades 3-5, 6-9, and in high school. New requirements included AYP measures for Title I schools only, which resulted in dual accountability systems for 28 states that were comprised of multiple performance indicators for Title I and non-Title I schools.⁴⁷ The dual accountability system made it difficult to measure and compare outcomes between schools based on different indicators, which also led to differentiated interpretations and uses of assessment results for Title I and non-Title I schools. In January 2002, the enactment of NCLB created a single accountability system imposing the same requirements on all schools. The same standards-based assessment is required to be administered to all schools in order to hold them accountable to the same content and performance standards.

Today, all schools are required to make AYP toward achieving NCLB's long-term goal; 100% of all children will demonstrate academic proficiency in reading and math by 2013-2014. The new national logo, "All Children Can Learn," is intended to refute past beliefs that only an elite group of students can learn.⁴⁸ States are now required to use a single accountability system to determine AYP for all schools based on year-to-year changes in the percentage of students who demonstrate proficiency on standards-based assessments. The method for determining AYP requires that each state establish baseline results to measure progress by comparing the baseline results to results obtained in subse-

46. E.L. Baker et al., *supra* n. 28, at 1-4.

47. M.E. Goertz & M.C. Duffy, *supra* n. 19, at 1-7.

48. L.A. Shepard, *supra* n. 7, at 4-14.

quent years. Unfortunately, this method of accountability creates concerns regarding the validity and reliability of interpretations and uses of results.

1. Concerns with Methods Used to Determine AYP

AYP is currently determined by rank ordering schools based on assessment results. Schools ranked toward the top are commonly described as effective and successful while those ranked toward the bottom are generally described as ineffective and in need of reform. A concern with using rank ordering to determine school effectiveness is that schools that are close in rank may be very close or very far apart in actual performance.⁴⁹ For example, when considering two elementary schools that are ranked 40th and 41st based on annual 4th grade reading results with mean results equal to 68 and 66.5 Normal Curve Equivalence (NCE), respectively, one generally assumes that the school ranked 40th is more effective at providing instruction that results in higher levels of academic achievement for students. This assumption may not be correct and more information is needed to clearly and accurately make that determination. If the added information showed that during the previous year the mean 3rd grade reading results for the same groups of students were 75 and 42 NCEs for the schools ranked 40th and 41st, respectively, then the only logical interpretation would be that the school ranked 41st was effective at improving academic achievement while the school ranked 40th was not. The school ranked 41st improved achievement results from 42 NCEs in 3rd grade to 66.5 NCEs in 4th grade. This group of students increased their average results by 24.5 NCEs while the school ranked 40th decreased their average results by 7 NCEs from one year to the next. This example further illustrates the importance of accounting for initial levels of academic ability in order to determine how effective a school is at improving academic achievement. Truly effective schools don't simply obtain high test scores, they demonstrate a consistent ability to adequately improve academic achievement for all students—that includes students with special needs.

In addition to rank ordering, other methods of accountability are used that also produce results that continue to be misinterpreted and misused.

49. C.W. Buckendahl et al., *District Accountability Without a State Assessment: A Proposed Model*, 21 *Educ. Measurement: Issues and Practice* 6, 6-16 (2002).

For example, a modified version of the Value-Added⁵⁰ method was developed and used in Arizona as MAP⁵¹ for schools. Results from this method are based on a comparison of a student's Stanford 9 results from one year to the next. Schools are described as effective when students demonstrate One-Year's-Growth (OYG) on the state's norm-referenced assessment, which simply requires that students maintain their stanine score from one year to the next. The problem with this method is that students who simply maintain below average stanines increase the likelihood that they will fail the required standards-based assessments, especially as they advance onto high school where they will be required to pass the state assessment to be eligible for graduation. The continued use of this methodology in Arizona Learns⁵² suggests that those who continue to promote value in this method have failed to recognize and understand that OYG is not AYP for students who are at-risk of academic failure because they will need to achieve more than OYG in order to increase the probability that they will pass the required high school graduation assessment.⁵³ Despite the limitations of the MAP, this method continues to be described as fair and accurate and is used to reward some schools for failing to adequately improve academic achievement for at-risk students through the state's accountability system.

2. Problems with AYP Due to Timelines and Operational Definitions

States have confronted a number of challenges with NCLB requirements. For example, the amount of time that was provided to account for AYP made it difficult to develop and use valid and reliable assessments. The problem was that NCLB required that states implement an accountability system to determine AYP immediately, which made it impossible to develop a method of accountability based on empirical evidence that clarified and justified the interpretations, uses, and consequences of results. The empirical evidence needed to clarify and justify the inter-

50. Tennessee's Value-Added Assessment System was developed by William Sanders and used as a measure of teacher effectiveness. A similar method was adopted and modified in Arizona and is referred to as Measures of Academic Progress.

51. For a definition of Measures of Academic Progress (MAP), see <http://www.ade.az.gov/azlearns/PuttingPieces/01ABCs.doc>.

52. Arizona Learns is Arizona's state accountability system used to label schools. Labels include Underperforming, Performing, Highly Performing, and Excelling.

53. A. Duran, *Evaluating the Interpretations, Uses, and Consequences of Arizona's School Achievement Profile Results*, (unpublished Ph.D. dissertation, Univ. of Ariz. 2003).

pretations and uses of results was missing because NCLB enacted in 2002 required that AYP be determined and reported that same year, which meant that the process of collecting data to conduct empirical studies had not begun because standards and standards-based assessments were in the initial stages of development. In addition, NCLB required that standards-based reading and mathematics assessments be developed for grades 3 – 8 and high school by 2007 to determine AYP, which suggests that the data that are needed to conduct empirical studies on AYP are nonexistent and will remain so until the standards-based assessments that are currently being developed are fully operational.

The lack of empirical data needed to conduct empirical studies on AYP also makes it difficult to operationalize AYP. Before establishing criteria for expected growth, empirical research must be conducted to determine what is typical progress as well as what is truly obtainable progress for students under optimal school conditions. Currently, baseline data are not available to reflect typical yearly progress, let alone determine what is adequate.⁵⁴ The process of defining ‘adequate’ involves making difficult educational judgments about what is potentially attainable progress under reasonable assumptions.⁵⁵ Traditional measures of progress and standards referenced approaches share a limitation: Neither incorporates a gauge for adequacy of progress.⁵⁶

The operational definitions for critical accountability variables are a key piece of information that is needed to determine the purpose for accountability systems. For example, AYP is defined as “continuous and substantial yearly improvement toward achieving proficiency and advanced performance levels.” The vagueness of this definition makes it difficult to determine the purpose or the clear intent of AYP. In addition to vague definitions, it is difficult to assess student growth with testing programs that do not involve the types of equating and scaling procedures that are used in large-scale norm-referenced testing. As a result, it is also difficult to measure and monitor “continuous and substantial yearly improvement.”⁵⁷

AYP is based on school serving as the ‘unit of analysis’ as opposed to the class or student. AYP is determined by comparing the academic

54. R.L. Brennan, *supra* n. 4, at 6-18; R.D. Schwarz et al., *supra* n. 4, at 26-35.

55. R.L. Brennan, *supra* n. 4, at 6-18; E.H. Haertel, *supra* n. 4, at 5-9; R.D. Schwarz et al., *supra* n. 4, at 26-35.

56. R.D. Schwarz et al., *supra* n. 4, at 26-35.

57. R.L. Brennan, *supra* n. 4, at 6-18.

achievement results for different groups of students within a school as opposed to being a longitudinal measure of academic progress for individual students. Operationally defining and measuring AYP at the student level may do more for improving academic achievement for individual students, especially those who are at-risk of academic failure. Sufficient and quality prevention, intervention, and remediation efforts are most likely to occur for students who are at-risk of academic failure when they are also included as targets for improvement and success. They are more likely to become the targets for improvement and success when they become the “unit of analysis.”

3. *School Challenges and AYP*

In addition to challenges associated with operationally defining and measuring AYP, NCLB comes at a time when schools are also facing the many new challenges that are associated with changing clients and limited resources. Schools are now accountable for the learning of all students, which include an increasing number of disadvantaged students who, in many cases, are confronting multiple challenges (e.g., poverty, lack of nutrition, ELLs). Schools are also required to meet the needs of many new students who continue to enter the United States from other countries. It is not uncommon for older students who arrive into the United States to be placed into classrooms on the basis of their chronological age, which will create tremendous challenges for schools when students are not literate or have never attended a school in their country. Schools do not have adequate resources to successfully meet the needs of changing and challenging clients, in part, because the current system was never intended to assure successful learning of high standards curriculum by all students.⁵⁸ The lack of adequate resources and changing clients creates even greater concerns toward the interpretations, uses, and consequences that are associated with AYP.

Today, schools are confronting the challenges of public perceptions suggesting that they have not lived up to expectations. The power of public perceptions has also created a shift in the control of accountability systems, going from local to state jurisdiction.⁵⁹ For example, state accountability for federal funding is no longer based on results obtained from district assessments. The shift from district assessments to state

58. L.W. Lezotte, *Effective Schools Research a Proven Framework for Continuous School Improvement* (paper presented at the Model Schools Conference, Orlando FL).

59. C.W. Buckendahl et al., *supra* n. 49 at 6-16.

assessments for purposes of state accountability may have occurred because of major differences between district and state assessment results where students who performed well on district assessments were not able to demonstrate the same levels of performance on state assessments. Such outcomes could easily create a lack of public trust in a district's ability to clearly and accurately monitor academic achievement and progress through the use of their own assessments. Over time, changes in societal conditions and the increased public awareness about everyday school operations have resulted in changed expectations directed toward schools, which have further resulted in revised federal educational policies and requirements. The fact that change does occur without the use or support of empirical evidence also demonstrates the volatility of the entire accountability enterprise, as strong public perception and expectations is all that is needed to create changes in policy and requirements. The sanctions and consequences that are derived from the interpretations and uses of results are what establish the need to determine the accuracy of interpretations in order to clearly determine whether the intended uses are meaningful, useful, and justifiable.

D. Sanctions and Consequences

The use of sanctions and consequences is intended to create incentives for schools that do not make AYP to undergo reform efforts that produce significant, sustained, and continued improvements in the academic achievement of their students for fear of losing federal funding or losing authority to manage and operate their schools. Unfortunately, the prescribed sanctions and consequences are not based on empirical evidence that clarifies and justifies the effectiveness of their use, which creates concerns toward their validity. Once again, the process of validation is complex as researchers acknowledge the difficulty of interpretations for an observed score and the even greater difficulties associated with reaching consensus on the appropriate uses of test scores in applied context.⁶⁰ Recognizing differences in school context is a first step toward understanding the unique set of challenges that individual schools encounter.

1. School Context

An example is used to shed light on the significance of the concept of context. In the example, two types of schools—those likely to be identi-

60. M.T. Kane, *supra* n. 4 at 319-342.

fied as making AYP and those least likely to be identified as making AYP—are described to provide a contextual picture that may identify differences in the levels of challenges for each type of school—this can also be thought of as significant and meaningful differences in the difficulty levels that are associated with their unique challenges and responsibilities. Schools that are identified as not making AYP are commonly high-poverty schools where the majority of students enrolled are at-risk of academic failure and come from impoverished homes while affluent schools—where the majority of students come from social and economically advantaged homes—are most likely to make AYP. The challenges for high-poverty schools toward making AYP are likely to be immediate and large in scale while the challenges for affluent schools are likely to occur in later years in a relatively small scale because the vast majority of their students are academically proficient and have been since early in life.

2. Contextual Correlations

The duration and the magnitude of a school's challenges toward making AYP are determined, in part, by the method used to determine AYP. During the initial year of AYP implementation, each state was required to develop a threshold level of performance specifying the percentage of students who needed to demonstrate proficiency. The method was based on rank ordering schools through the use of assessment results. Schools ranked in the bottom 20% were immediately identified as not making AYP and were immediately subject to consequences. The fact that the majority of high-poverty schools were in the bottom 20% is not surprising because of the positive correlation between the social and economic status of student's homes and their academic achievement; high-poverty schools are more likely to be ranked at the bottom while affluent schools are more likely to be ranked at the top. The method used to determine AYP makes predicting school outcomes through the use of home income information about as reliable as predicting school outcomes from student achievement results.

3. Contextual Factors Define School Challenges

A comparison between high-poverty and affluent schools is provided to further demonstrate their contextual differences, which also create differences in the levels and degrees of challenges that are presented to

each type of school. A school's student population is generally comprised of children who live within the surrounding neighborhood. The level of family income commonly determines the neighborhoods that people can choose to live in. Affluent schools exist because the majority of families in the surrounding neighborhood earn high incomes that allow them to live in high priced homes while high-poverty schools exist because the majority of families in the surrounding neighborhood are poor and struggle to continually acquire the income that is needed to keep their home from month to month. For poor families, education may not be a priority because their primary concern is to find basic means for daily survival. Social and economic factors are often related to other factors that further define the nature and the magnitude of individual school challenges.

Students who are at-risk of academic failure and attend high-poverty schools are generally minority, tend to be poor, come from homes where parents do not speak English, generally have parents who have very little education and little or no knowledge of how the school system works, have cultural values and characteristics of their parents that are often at odds with the values and characteristics of the school and those who carry out its mission, and often find themselves in schools that are overburdened and not adequately equipped to meet their needs. Furthermore, students who are at-risk of academic failure often: Lack English proficiency; possess cultural differences; lack prior schooling; are mobile or homeless causing them to experience constant absenteeism; are placed in inappropriate classrooms that do not support their learning styles; and encounter stereotypes that have the capacity to destroy their confidence toward successfully achieving current and future goals.

Examples of stereotypes inflicted on minority students include, but are not limited to, preconceived notions of ability and motivation held by peers, teachers, and society. Such stereotypes tend to produce lowered expectations by teachers that continually impact educational opportunities and academic performance. In addition, stereotypes often result in minority students being placed into special education and remedial or vocational tracks as opposed to college-bound courses.⁶¹ To become successful in school, students at-risk of academic failure must learn to con-

61. M. Suarez-Orozco et al., *Cultural, Educational, and Legal Perspectives on Immigration: Implications of School Reform*, in J.P. Heubert, *Law & School Reform: Six Strategies for Promoting Educational Equity* 165, 165-179 (J.P. Heubert ed., Yale University Press 1999).

front and overcome tremendous challenges that enable them to remove barriers while acquiring the high levels of confidence, discipline, and motivation that are needed to acquire the academic skills and knowledge that ultimately lead to continued academic and future success. These students will also need to find at least one adult who is sincerely committed to continually assisting and directing them in their continued pursuit of learning and success. Academic and future success is a major undertaking for students whose homes do not provide the social, emotional, educational, and economic support that are needed.

4. Contextual Resource Differences

High poverty and affluent schools also differ in availability of resources. Research suggests that disparities in educational resources and opportunities that are provided to students generally mirror disparities in their educational attainment.⁶² Although federal funds are provided to high poverty schools, it is not clear that the allocation is adequate for improving academic achievement. If federal fund allocations are not adequate for meeting the needs of students, then it is difficult to justify or explain the use of sanctions and consequences on high-poverty schools. Requiring use of an accountability system to impose sanctions and consequences on high poverty schools while failing to provide adequate funding to meet the needs of all their students is simply a recipe for failure and disaster. School reform efforts that prove to be effective will most likely also demonstrate a successful acquisition of sufficient and quality resources used to provide enriched and effective learning experiences and opportunities for students, which further enhance their continued academic progress and success.

If fairness and validity are truly important concepts in accountability, then sanctions and consequences must not be imposed on schools that do not receive adequate funding. A lack of adequate funding restricts schools from acquiring sufficient and quality resources to provide enriched and effective educational experiences and opportunities for all students. Inadequate funding also has a negative effect on the small number of high achieving students who attend high-poverty schools because they are most likely to be neglected. This occurs because the

62. M.S. MsUSIC (1999), *School Finance*, in J.P. Heubert, *Law & School Reform: Six Strategies for Promoting Educational Equity* 93, 93-100 (J.P. Heubert ed., Yale University Press 1999).

limited school resources that are available are targeted toward meeting the needs of students with the greatest needs. The lack of resources that are needed to better meet the needs of high achieving students who attend high-poverty schools increases the likelihood that their academic performance will regress toward their school's level of academic achievement.

5. Contextual Validation of School Results and Their Interpretations

As stated previously, the validation process requires an accurate interpretation of observed scores in order to reach consensus on the appropriate uses of test scores in applied context. If students who attend affluent schools are just as likely to pass the end of year state assessment at the beginning of the same school year, then one must reconsider the interpretations of results that suggest high levels of school effectiveness. Students who attend affluent schools generally come from homes where parents have generally attained higher levels of education that have enabled them to provide enriched learning experiences and opportunities for their children as early as their prenatal years. On the other hand, students who attend high-poverty schools and are at-risk of academic failure generally come from disadvantaged homes where they have not been afforded the luxuries of having had received enriched learning experiences and opportunities in their homes from the time that they were born. It is also likely that their parents were not able to afford the luxury of their prenatal care. Birth generally marks the beginning of life's challenges for children who attend high-poverty schools and continue to be identified as at-risk of academic failure. They will undoubtedly begin their academic careers possessing and exhibiting deficiencies and needs in many critical areas of life, including education.

An interpretation of results suggesting that affluent schools are the primary cause of high academic achievement requires further inquiry. The validity of interpretations and uses of AYP results is questionable when school outcomes, home income, and student achievement are highly correlated with one another. It may be that home income factors are the most direct causes of academic achievement. If accountability systems determine school effectiveness on the basis of academic achievement results without the inclusion of added contextual variables, then one must question the validity of the results because school populations are not created through a process of random selection and random assignment. The contextual reality may be that the challenges that

are encountered by high poverty schools are more difficult and greater in number than the ones encountered by affluent schools.

6. The Context of Opportunity-to-Learn

Opportunity-to-Learn is another contextual difference that exists between high-poverty and affluent schools, which is based on whether or not a school provides all students with sufficient and quality time and resources that enable them to learn the required skills and knowledge.⁶³ The fact is that high poverty schools are responsible for providing the same levels of OTL within the same window of instructional time allotted and required for all schools. This is a difficult task since providing sufficient instructional time means providing more time for students who need more, which will require extended time beyond a normal day and a greater number of quality and effective resources. If this does not occur, then it is highly unlikely that students who are at-risk of academic failure will be able to make AYP toward achieving proficient levels of achievement. The degree to which sufficient and quality OTL may or may not be provided for all students may also be a measure of the degree to which OTL is not equally provided from one school to another. One must recognize that simply taking away instructional time from higher achieving students in order to provide more instructional time for students who are at-risk of academic failure will generally result in an insufficient amount of instructional time for all students. This method of providing sufficient and quality resources does not constitute OTL.

V. CONCLUSION

NCLB creates the opportunity to develop and implement valid and reliable accountability systems that clearly and accurately identify effective schools. AYP measures must be based on the amount of annual progress made by students as opposed to simply being based on a rank ordering of end of year school results at benchmark grades, which comprise of different groups of students. To be identified as effective, a

63. R.M. Jaeger, *Certification of Student Competence*, in R. L. Linn, *Educ. Measurement* 485, 485-514 (R.L. Linn ed., 3d ed., American Council on Education/Macmillan 1989); P. Weckstein (1999), *School Reform and Enforceable Rights to Quality Education*, in J.P. Heubert, *Law & School Reform: Six Strategies for Promoting Educational Equity* 309 (J.P. Heubert ed., Yale University Press 1999).

school must demonstrate the ability to improve and accelerate the academic achievement of low achieving students while maintaining or improving the academic achievement of students with average and above average ability. To be identified as Excelling, a school must demonstrate these levels of effectiveness for a diverse student body population where a significant number of students have complicated needs.

It is suggested that the challenges presented by NCLB are difficult but worthy of confrontation in order to develop and establish educational systems where all students receive equitable and excellent learning experiences and opportunities at every school, grade level, and classroom. As one moves forward one may discover that the initial pursuit of the most important goals in life is generally difficult but the process is simplified as we continue to learn and improve. One must also recognize that validation is a process that requires continual refinement of ideas, practices, and policies in order to better meet the needs of our most precious resources, our students. We owe it to our students to develop an accountability system that ensures their success in school and in life. If approached and used properly, NCLB can become the vehicle that transforms ordinary classrooms into extraordinary learning environments where all students are able to discover and reach their full potential in order to successfully pursue and achieve their life-long goals and ambitions.

The process of continually validating existing or evolving interpretations and uses of results is essential for systemic improvement. Validity must serve as an integrated evaluative judgment of the degree that empirical evidence and theories support the adequacy and appropriateness of inferences and actions that are based on test scores. Existing evidence is enhanced over time by new findings and changes in social conditions that serve to create changes in the interpretations, uses, and consequences associated with test scores.⁶⁴ Educational accountability systems must be used to accurately identify schools that have clearly demonstrated improved changes leading to positive outcomes. In addition, accountability systems must accurately identify schools in need of improvement so that they are able to receive the type of support that is needed for them to improve and to succeed.

64. S. Messick, *supra* n. 4, at 1-104.

A vertical bar on the left side of the page, consisting of a series of horizontal segments in shades of yellow and orange, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Factors to Consider When Evaluating School
Accountability Results

SOURCE: J Law Educ 34 no1 Ja 2005

WN: 0500102292004

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher:
<http://www.law.sc.edu/jlawedhp.htm>

Copyright 1982-2005 The H.W. Wilson Company. All rights reserved.