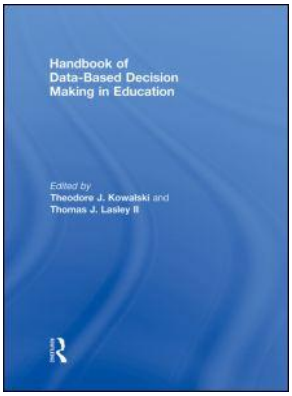

Handbook of Data-Based Decision Making in Education

Theodore J. Kowalski, Thomas J. Lasley

The No Child Left Behind Act

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203888803.ch5
Timothy R. Konold, James M. Kauffman
Published online on: 13 Oct 2008

Handbook of Data-Based Decision Making in Education

Edited by
Theodore J. Kowalski and
Thomas J. Lasley II

First published 2009
by Routledge
270 Madison Ave, New York, NY 10016

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

This edition published in the Taylor & Francis e-Library, 2008.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
Handbook of data-based decision making in education / Theodore J. Kowalski & Thomas J. Lasley II, editors.
p. cm.
Includes bibliographic references and index.
1. School management and organization–Decision making–Handbooks, manuals, etc. I. Kowalski, Theodor 1943– II. Lasley II, Thomas J. 1947–
LB2805 .H2862 2008
371.2 22

ISBN 0-203-88880-4 Master e-book ISBN

ISBN10: 0–415–96503–9 (hbk)
ISBN10: 0–415–96504–7 (pbk)
ISBN10: 0–203–88880–4 (ebk)

ISBN13: 978–0–415–96503–3 (hbk)
ISBN13: 978–0–415–96504–0 (pbk)
ISBN13: 978–0–203–88880–3 (ebk)

5

The No Child Left Behind Act

Making Decisions Without Data or Other Reality Checks

Timothy R. Konold and James M. Kauffman
University of Virginia

Making Decisions Without Data or Other Reality Checks

A very strange thing happens to many people's thinking when education is the topic: they do not make use of reality checks, including knowledge of basic mathematics and fundamental statistical concepts, critical thinking, and data. They seem to assume that education operates in an alternate universe in which realities do not apply at all or may safely be assumed to be trivial or irrelevant to the task of teaching students. They seem not to make logical connections between words and their referents, not to employ basic concepts, not to demand reliable data, and not to engage in logical inductive or deductive thinking, all of which they rightfully expect of school children. Yet their discourse about education is taken seriously, and the result is what one would expect: failure. Given this level of discourse about education, it is no surprise to find the conclusion that "our so-called system of education is far less well planned and executed than our system of highways and of mail delivery" (Shattuck, 1999, p. 34). The actual improvement of education awaits the application by everyone whose discussion is taken seriously of data-based, reality-based thinking skills (Kauffman, 2002).

We realize that life is full of absurdities, but we think that rhetoric about education should not be among them. When someone suggests that all children will be able to perform at ___ (>0) level, that all children will succeed, that all students will be proficient, or that no child will be left behind, he or she is contributing to unhelpful silly talk about schools and schooling. Some individuals may argue that the No Child Left Behind Act (NCLB) at least focuses attention on education and motivates educators to work harder to make sure that all students achieve what they can. But NCLB is a prime example of absurd education policy that is divorced from data and reality checks about the meaning of data. (We assume that the basic provisions of NCLB are described in other chapters in this volume.)

Policies like NCLB, though conceptually flawed, may produce some benefits for some students, perhaps even for many. In fact, just about any policy enacted with good intentions, regardless of how conceptually vacuous, will benefit *some* students. However, policies like NCLB—based on irrational premises and ignoring important realities—are not only doomed to collapse, but in the longer term hurt the cause of education more than they help. We are not the only ones to note the irrationality of

NCLB and the blithe avoidance of realities by those who enacted it. "Oxymoron" is an accurate descriptor of the universal proficiency demanded by NCLB (Rothstein, Jacobsen, & Wilder, 2006).

Unfortunately, NCLB is not the only poorly conceived education policy. It is only a convenient and popular example. Senseless statements about education and policies to match them are common but incompatible with education's goals. "Especially in the case of education, which at its most fundamental level is helping people make sense of things, senseless rhetoric runs against the grain of what one is trying to accomplish" (Kauffman, 2002, p. 284). *Senseless rhetoric* describes much of what has been written and said about education by many educators, captains of industry, politicians, newspaper columnists, and others. Rhetoric may safely be judged senseless when it ignores any reality.

Most people realize that when they ignore data and other realities they are courting disaster. In fact, when someone ignores realities most observers conclude that he or she is mentally ill or at least is engaging in very dangerous behavior. We try to teach our progeny and our students about realities so that they do not harm themselves or others. If they ignore realities, we apply restrictions and warn them about the dangers of assuming that reality does not affect them. We want our children and our students to talk and act as if they are in a real world in which some things are universally, unalterably true, whether anyone likes it or not and where the reliability of data is important.

However, too often educators and others seem to assume that data and other realities are distractions, unimportant or avoidable in their proposals to reshape, restructure, or improve education. Their assumption may be based on one or more false premise. Sometimes, it is based on the philosophical proposition that what we consider realities are socially constructed by one's culture, and therefore not universal. Sometimes it reflects ignorance, misunderstanding, or presumed political finesse. Sometimes it is expressed as a desire to "think outside the box," as if the "box" of realities is an artificial constraint that needlessly inhibits creative thinking. It may sometimes reflect extreme frustration in trying to change education policy, such that an unrealistic policy may be supported as the best one can hope for—at least a beginning of change in the right direction. Or it may reflect the assumption that moral decisions can be made in the absence of reliable evidence or the misrepresentation of data to fit one's purposes (see Brantlinger, 2006; Gallagher, 2004; Gartner & Lipsky, 1989; Kohn, 1999, for examples of evading or denying realities while calling for reform). But regardless of *why* anyone thinks and talks as if realities do not exist in education or that data can be interpreted willy-nilly for one's benefit, this gambit has adverse consequences.

Also too often, any statement about education or any call for reform, regardless of its failure to conform to the real world, goes unchallenged. Some seem to assume that every statement or suggestion about education should be taken seriously. The consequences of this assumption are dire: actual improvement of education is stymied while possibilities that are only imaginary are pursued. Long ago, we learned that some things (e.g., turning lead to gold) are flatly impossible. For many things, proof is demanded in the scientific tradition before someone is allowed to market an idea or product. Nevertheless, as a society we have not only allowed but supported and

sometimes even enshrined in law preposterous propositions about education. Silly talk—pretense—has too often been allowed to share a place at the table with serious talk about education. This produces babble, not constructive conversation.

In many ways, education *is* rocket science—complex, difficult to understand, and disastrous when one seemingly small detail is forgotten or one fundamental reality is overlooked. Space shuttle disasters occurred because little realities were neglected; getting things mostly right but making one critical assumption incorrectly can result in disaster. Getting things mostly right in education is not good enough either. In education, we too often forget or ignore realities that may be considered inconvenient but that bring about an inevitable policy failure when they are assumed to be trivial.

Education as an enterprise is supposedly designed to help students separate fact from fiction and deal productively with the confusing world in which we live. Teaching based on the premise that fact cannot be distinguished from fiction will always be derelict. So will talk of education that does not frankly acknowledge realities—and reject as frivolous—the fictions to which people cling implicitly when their thoughts and actions reflect ignorance of or disregard for realities.

Most people understand how projects in the physical or social sciences are doomed ultimately to failure by ignorance of or disregard for realities. For example, no one is allowed to construct and sell airplanes if he or she has ignored any reality of gravity or aerodynamics in the process. No view of criminality that ignores the necessity of negative consequences for criminal behavior is taken seriously. Talk of education that does not recognize important realities should not be accepted, simply because any proposal for reform, legislation, or regulation that does not start from the premise that realities must be incorporated into the plan is doomed to failure. Reality-evading proposals may be appealing to the masses, may be sold to the unsuspecting, and may seem initially to be successful. Ultimately, they crash and burn, as does any reality-evading proposal in the natural sciences. NCLB is a case in point of education policy doomed to failure because it ignores realities. It is decidedly not even logical or reality-based, much less evidence-based.

NCLB, Measurement, and Distributions

The improvement of education depends on attention to more realities than those involving measurement (see Kauffman & Konold, 2007), but the realities on which we focus in this chapter involve what we know about distributions of data. In particular, we discuss the necessity of measurement and of attending to all of the mathematical realities of statistical distributions.

Indispensability of Measurement

Measurement is essential to any science. Without measurement, no scientific venture can be seen as legitimate, much less succeed. In education, we must have measurement to know how our students are doing. Otherwise, we have no way of judging whether teaching is having any effects on our students. Teaching without wanting to

know how students are doing as a result of our teaching is what Sasso (2001) calls "willful ignorance."

Although teaching may have its highly individualistic features, it is also based on—and only evaluated legitimately by—methods that have a scientific base in reliable data. The call for scientifically validated methods of teaching is not misguided. However, one cannot reasonably call for scientifically validated methods of instruction and then ignore data on which science is based. The fact that educators, politicians, or others have sometimes embraced the notion of scientifically validated instruction while ignoring realities (e.g., data on which scientific principles are based), merely illustrates the way in which discussions of education and proposals for its reform can go awry.

Some may oppose measurement, including classification and ranking, yet deny that accountability implies measurement (e.g., "Obviously, I am not eschewing accountability for what we do. I am breaking the particular equation of accountability = measuring/ranking," Heshusius, 2004b, p. 295). But people can write explicitly self-contradictory statements, make a claim and then refute it in the next sentence, make explicit statements that are the opposite of what is implied in another of their statements, and write or speak gibberish that others of similar bent defend as understandable or even describe as insightful (see Kauffman, 2002; Kauffman & Sasso, 2006a, 2006b; Mostert, Kavale & Kauffman, 2008).

Measurement in Education Requires Testing

Any type of test is a way of checking to see whether students have acquired particular skills. Tests may be essays, performances, portfolios, or those types most often criticized—multiple item, more objectively scored tests that may be true/false statements, fill-ins, or multiple choice. Tests are sometimes misused or abused. More often, they are simply misunderstood. The most dangerous misunderstanding is that testing can be avoided. Those who criticize NCLB because it requires standardized testing are on the wrong track (see Kohn (2000, 2001) for examples of criticizing standardized testing). NCLB is misguided policy, not because it requires standardized testing but because it is based on misunderstanding of the statistical properties of tests and on warped interpretations of the meanings of test scores.

All types of tests have advantages and disadvantages. None is perfect, but some are very good at giving us important information. Of all tests, we must inquire about what is being assessed, how the knowledge or skill of interest is being measured, who comprises the comparison group, and what the obtained score means. Tests may be criticized on any of these grounds—that they measure the wrong thing or something trivial; that they are of questionable reliability or validity; that they result in inappropriate or invidious comparisons; or that the obtained score is meaningless or misleading. Standardized tests have been the objects of scorn primarily because critics do not understand what they are designed to do and why they are important. Standardized tests have also been criticized because the results have often been used inappropriately. But workers, not their tools, should be held accountable for shoddy work.

Test developers and measurement specialists are well aware of the limits of tests. Tests serve as tools, but as the Standards for Educational and Psychological Testing of the American Educational Research Association and American Psychological Association (1999) remind us, decisions about individuals or groups should not be made on the basis of a single measure. Recently, editors of several measurement journals crafted a joint letter to various NCLB policy makers highlighting the dangers of making high-stake decisions on the basis of a single measure. Among the points made were the facts that a single measure does not have the psychometric quality to serve as the sole basis for making high stakes decisions, and that such decisions should be based on multiple *sources* of evidence that are reflective of student achievement (Fitzpatrick, 2007).

Aside from education, most people understand the importance of testing. Most believe it is important that products be tested before they are marketed. And most desire services from professionals who have passed tests of their knowledge and skill. Testing is required for quality control. Quality control in education is needed, but children are not products that can be manufactured to a uniform tolerance. Other aspects of the measurement of human beings preclude treating children as if they can be discarded if they do not measure up to a given criterion, not the least of which is the reality that measurement of human beings always results in what Gould (1996) refers to as a "full house"—a full distribution (in the present discussion, of test scores or other performances), any part of which is ignored only at the cost of denying reality.

Measurement is Required for Identifying Failure and Success

The words *failure* and *success* are nonsensical in the absence of measurement. Failure and success for an individual are always defined by that student's performing or not performing some expected task or by a student's reaching or not reaching a given criterion. To suppose that failure is impossible or that a student will be judged successful regardless of performance is to deny the meaning of the terms. Certainly, saying that all fish are large is nonsensical; saying that all those that do not pass through a net are large may make sense. Likewise, to say that all children are successful (or successful at ___) is nonsensical, although all who can demonstrate a certain level of proficiency in a particular activity may be judged successful.

One legacy of bad thinking about education is the assumption that failure at one thing means failure in general. The solution to this problem is not to try to pretend that failure can be avoided for all students but to recognize that failure by a given measure signifies only failure on what was measured. True, we "make" failure by our measurement, and we cannot do otherwise unless we are to draw no lines or set no expectations for accomplishment for any purpose or for any age. Any measurement that produces only successes is bogus—it does not reflect the real world. Nevertheless, failure by one criterion does not imply failure by every criterion.

Measurement is Necessary for Accountability

Regardless of whether measurement is a very subjective judgment of a more objective judgment such as a certain score on a standardized test, measurement is required if educators are going to be held accountable. In fact, accountability in the absence of measurement is a non-sequitur—it does not follow, and indeed cannot follow, that accountability has been demonstrated in the absence of measurement of either student performance or teacher behavior. Those who call for accountability and moral judgment but condemn testing and measurement (e.g., Heshusius, 2004a, 2004b) are not describing something that comports with the actual world of accountability—except that a person may be empowered to judge that someone has or has not been "accountable" or "successful" by undisclosed criteria, a tactic of misanthropes.

Measurement Allows Us to Identify and Work on Gaps

Most educators and observers of education are aware of gaps in achievement, knowledge, or performance (regardless of what term one uses to describe evidence of learning) among individuals' abilities and among various groups (which may be categories defined by ethnicity, gender, age, class, or disability).

Educators may wish to widen the gap between the means for those who are gifted and those who are not, but we want to narrow or close most gaps between the means of groups. Without measurement, it is impossible even to make the case that gaps exist, and it is impossible also to know whether gaps are widening or narrowing. Thus, any suggestion that achievement gaps, disability, or diversity can be identified without measurement (e.g., Heshusius, 2004a, 2004b) is nonsense, as is denial that measurement always reveals differences that may be interpreted as gaps. Some gaps may be impossible to close by any legitimate means, but others can and should be narrowed or closed altogether.

One of the problems with NCLB is that it ignores realities about the nature of gaps and what is required to close gaps of various kinds. It does not seem to recognize that

(a) there is a mean of means (i.e., that means, as well as scores for individuals, have a distribution),
(b) some of the variance in achievement scores is due to factors other than teaching,
(c) the mathematical properties of continuous distributions apply in all cases, and
(d) besides mean differences, other differences in distributions reveal important gaps.

The failure of NCLB to recognize realities regarding all of the various gaps in distributions of achievement scores is a fatal flaw—as certain to result in disastrous consequences as the failure of the National Aeronautics and Space Administration to take into consideration the effects of cold temperatures on booster rockets' O-rings and the damage done by foam insulation striking surfaces at high velocity.

NCLB's Lack of Attention to the Realities of Individual Differences and Distributions

Supporters of NCLB argue that schools have not been successful in educating all students. In fact, the rationale for NCLB is that schools do not "work" or are "failures" (see Popham (2004) for descriptions of how NCLB defines "failing" schools; criticism of public schools as failures and calls for reform so that they "work" has a long history, as Cuban (1990) recounts). Saying that schools are "failing" begs the question of what we should expect to see if schools or any of the programs offered therein are "working" (Kauffman, 1990). Would children's test score distributions be different, and if so how? Would measures of central tendency (e.g., mean, median) be higher? Would we expect to see all students obtaining the same score on these measures with little or no tolerance for any variation in expected student outcomes? Is it really possible to raise the central tendency of test scores and at the same time reduce the variation in scores? These questions are central to resolving whether the goals of NCLB are attainable even if public schools are really good.

Individual differences are a reality. People differ on so many factors (e.g., interests, attitudes, motivation) that it is difficult to identify *in*variant characteristics. Student achievement is no exception. To illustrate, standardized achievement tests are typically normed so that the mean of the raw score distribution is set at 100, and the standard deviation is 15. For decades, the resulting scores have been shown to be approximately normally distributed when tests are administered to large populations of students. The reality of such distributions is that 50% of the students measured will have scores below the mean, and 50% will have scores above the mean. In fact, 68% of the students will have scores that range in value from 85 to 115, or a 30 point standard score difference; and 95% of the tested population will have scores ranging in value from 70 to 130, or a 60 point standard score difference. Children show individual achievement differences, and these differences were pervasive even before we started measuring them.

Some individual differences are important for instruction, but others are not (Kauffman, Conroy, Gardner, & Oswald, in press). The most important differences for instruction are differences in prior achievement (Engelmann, 1997). Some students are simply better equipped to accumulate knowledge at a faster rate than are others, and the uniform application of programs across students of different ability levels does little to help groups on either side (of the central tendency) of the achievement continuum—though it may serve to reduce the existing variability in student achievement. These ideas are in contrast with NCLB initiatives that assume a one-size-fits-all model (Lawrence, 2006). Moreover, individual student variation is likely to be masked by the currently adopted school level aggregates for classification (i.e., AYP vs. non-AYP) of accountability based on a single school-wide measure. For example, in a recent analysis of schools meeting AYP, benchmarks were found to be comprised of both below-average students demonstrating very little progress and above-average students showing material gains. Other schools, however, also meeting AYP benchmarks, were found to contain below-average students making positive gains with above-average students showing little progress (Choi, Seltzer, Herman, & Yamashiro, 2007).

Because of the punitive consequences of NCLB for schools that do not "measure up," administrators feel threatened by the looming prospect of losing resources when struggling students fail to reach proficiency. As a result, greater emphasis is placed on those struggling students who are relatively close to but below the expected standard, with fewer resources being made available for gifted education (Winerip, 2006) or those with cognitive limitations who obviously will not meet the standard. In the end, this may well result in distributions of achievements that show smaller differences between children at opposite ends of the achievement scale. However, these are not the gaps that we wish to narrow. This unfortunate consequence will also contribute to leaving students behind in a more important sense than their being lower than a population mean (i.e., leaving them far below their potential). That is, students holding the greatest academic potential for success in achieving domestic and international advances are likely to be "left behind" in that they are unchallenged and do not achieve what they could (Gentry, 2006; Goodkin, 2005).

Most Educational Variables are Continuous in Nature

Some educational measurement consists only of categories (e.g., male–female; did/did not meet a performance criterion) or ranks (e.g., first in class, second-to-last to finish). However, most measurements of educational performance (e.g., most standardized achievement tests) produce a score distribution that lies atop a continuous distribution of outcomes. As Kauffman (2002) and Kauffman and Hallahan (2005) point out, continuous distributions are those in which what is measured varies from a little to a lot with fine gradations or increments being possible. There are no natural, inherent, or obvious breaks in a continuous distribution. Height and weight are examples of continuous distributions, as is rate or speed.

Whenever human performance (as well as many other physical attributes of humans and many things in the material world) is measured, the results of measurement produce a discrete distribution. This occurs because, regardless of the degree of precision inherent in the measurement tool (tenths, hundredths, thousandths), there will be a break in the scale when moving from one of these measured points to another. Often, however, the data resemble what has come to be known as a "normal" distribution, in that a graph of the scores approximates the symmetry of a bell by modeling the underlying continuous scale that is inherent to the variable being measured (hence, "bell-shaped curve" or "bell curve"). Not all distributions or curves are "normal" or symmetrical; some are lopsided or skewed. Nevertheless, all continuous distributions have immutable properties.

The immutable properties of continuous distributions are referred to by psychometricians as "moments," of which there are four: central tendency, variability, skewness, and kurtosis. These are well-established realities in spite of attempts to cast them as mere human inventions (see Gallagher, 2006) or to ignore their implications for closing gaps in performance (see Kauffman, 2005). These distributions and immutable qualities apply to all groups—including all groups of averages for schools, districts, states, nations, and subgroups of students. They are as real and as useful as many other human inventions, including houses, arithmetic, roads, laws, and

languages. Moreover, criteria that may be used to categorize individuals (e.g., did or did not meet a criterion for reading ability—in short, all "criterion-referenced" tests and judgments of grade-level performance) are actually derived from the measurement of continuous distributions of performance; the criterion is based on a continuous distribution.

Philosophical objections and observations that the smoothed curve depicting a continuous distribution is actually drawn from a histogram (a bar graph) or that the first such curve was drawn from estimates of measurement error (Gallagher, 2006) do not mean that continuous distributions are not real. Assuming that continuous distributions can be ignored, changed, or violated at will because they are educational or psychological invites disaster.

The only way to consider no student or group (or no school, district, or state) "behind" or "low-performing" is to make questionable comparisons or comparisons that at least partially obscure the truth for the unsuspecting observer. For example, one can compare a given score or average to a distribution obtained long ago. Although it may be the case that we could have education so good that fewer than 25% of students score below the 25th percentile on a test given 25 years ago (or even on a more recently administered test), it is impossible to have fewer than 25% of students scoring below the 25th percentile on the test their group has taken for the purpose of describing the distribution. Thus, questions about the comparison group and the comparison distribution are always pertinent. But of even greater importance for the present discussion is the fact that there will be—always, every single time—a distribution, including a portion below average, a bottom quartile, a lowest score, and, if Gould's (1996) full house is considered, those whose disabilities are profound and whose scores are zero. NCLB blithely disregards this reality and is, consequently, doomed to disgrace as education policy.

No Measurement is Error-Free

Regardless of the type of measurement employed (whether categorical, ranking, or continuous; whether personal judgment of performance or a standardized multiple-choice test), it is not absolute or perfect. This is true also in the physical world; our measurements, even the most precise ones in physics, have a margin of error. Thus, measurements are always estimates, although some are more reliable, more valid, more precise, or more accurate (i.e., less variable) than are others.

Generally speaking, the less the margin of error, the more desirable the measurement. Those seeking to improve a particular field of study are always desirous of more precise and less variable measures. It is wise to recognize two realities about the measurements we use in education: (a) they are often highly useful and (b) they have a margin of error. The fact that educational measurement contains error is not justification for ignoring or refusing to use it. Error just needs to be recognized and taken into account in educational decisions based on measurement. Measurements that are known to have a margin of error (e.g., speedometers, thermometers, test scores of various kinds) are nonetheless useful. Margins of error, possible sources of error, and judgments based on measurement must simply be recognized for the

realities that they are. Measurement may be essential, but it does not preclude judgment. Yet in many assessments of progress related to NCLB, there is no reporting of the standard error of measure, so readers have no way of judging whether the reported differences in means are chance variations or statistically significant changes.

Yearly Measurement is Useless for Preventing Failure

Yearly measurement, like that demanded by NCLB for assessing progress, has its place in telling us how groups and individuals are doing compared to a standard. However, such measurement is useless for avoiding failure. Prevention of failure requires anticipating it or at least catching it in its incipient stages, not pointing it out long after the fact (see Kauffman, 2002, 2003, 2004). The kind of testing demanded by NCLB can tell us that something went wrong (or right), but it cannot tell us precisely *what* went right or wrong or even *when* it occurred, and it is always too late to avert disaster for an individual. The kind of measurement that is useful for averting prolonged failure is curriculum-based—frequent, if not daily, assessment by the teacher of a student's progress in the curriculum.

Moreover, no one has invented a measurement device that is reliable and valid, yet produces no variance in what is measured (probably such a device cannot be invented, as it contradicts what we know about the achievement construct we are measuring). And when a designated group of students is measured, it is simply a reality that there will be a lower or bottom part of the distribution, including those who are lowest or last, some of whom will be judged to have failed. This is why setting a universal standard for performance, as suggested by NCLB, is unrealistic—unless *universal* is taken to mean *all who do not fail*, in which case it is clearly nonsensical.

In addition, policy makers in NCLB seem to assume that proficiency holds the same meaning across tests and states (Fuller, Wright, Gesicki, & Kang, 2007). The reality, however, is that definitions of AYP differ from state to state because states use tests of different difficulty, different standards of proficiency, and measurements of different curricula. Thus, making realistic assessments of achievement gains as a nation is difficult at best. The flexibility afforded to states in setting their own performance standards distorts a common understanding of "proficiency." For example, King (2007) reported that 89% of fourth-grade students passed the 2005 state-sponsored reading test in Mississippi. By contrast, Massachusetts could only claim a reading pass rate of 50% among their fourth graders in the same year. These results would likely suggest that students in Mississippi were receiving better instruction—that is, until one considers how these groups performed on the National Assessment of Educational Progress (NAEP). Results on this common, across-state, measure indicated reading pass rates of 18% and 44% in favor of Massachusetts fourth graders. In general, analyses of NAEP data reveal material state-to-state levels of variation in terms of progress trends, but within-state trajectories of growth are fairly stable across different grades and content areas (Schafer, Liu, & Wang, 2007).

Statistical Data Other than the Mean are Critically Important

Perhaps the most disappointing thing about NCLB and many of its discussants is failure to consider statistical concepts other than central tendency. Gaps between groups in performance are apparently conceptualized only as differences in means; *the achievement gap* is an apparent reference to the difference in test score means for African Americans and Caucasians, although many gaps of many kinds could be discussed.

Typically, mean scores for groups are the target of concern, likely reflecting a failure to consider the best index of central tendency. The failure to consider variance and the nature of the dispersion of scores—standard deviation, standard error, kurtosis, and skew, for example—are more serious concerns. Means can lead people to make silly and invidious comparisons of groups when variance is ignored (see Gladwell, 1997). Moreover, statistically significant mean differences between groups can be of no practical significance when very large samples are compared, and significant mean differences may occur in the absence of any significant differences among groups in very high or very low performers.

Although researchers have applied sound methodological tools that capture student and school-level variations to the investigation of student growth (Choi et al., 2007; Schafer et al., 2007), virtually all discussion of NCLB in the popular press ignores these statistical realities and takes an extraordinarily simplistic view of critical issues about measurement of educational performance—so simplistic that it insults the intelligence of many who read or hear it. An issue of considerable importance in evaluating the effects of education is the dispersion of scores—the shape and nature of the distribution of scores. For example, would we see increased variance as an indication of progress or regression? Moreover, in evaluating the differences between means, the absolute difference between them is meaningless without knowledge of the variances of the samples (or populations) being compared. Yes, for the most part, academic discussion of growth or progress related to NCLB ignores or makes marginal use of estimates of variance (e.g., "A vote for 'No Child'," 2007; Fuller et al., 2007). We are not suggesting that educational researchers such as Fuller et al. (2007) or government officials are incapable of understanding distributions and their importance, but the fact remains that most of the discourse about NCLB has not included statistical concepts other than mean differences. In fact, issues of statistical distribution were totally ignored by both a 2007 *Washington Post* editorial ("Save School Standards," 2007) and every letter writer in the September 10 issue of that newspaper, including the U.S. Secretary of Education and the President of the National Education Association.

Regardless of the origin of the intellectually impoverished view of education reflected by NCLB and most of the discussion of it and its effects, the level of discourse about this law has not been raised by legislative debate, administrative argument, or even by most academic dialogue. In our opinion, it is time for people in academic and administrative roles to insist that the level of discourse be more firmly grounded in the realities of data, including the mathematical realities of score distributions that give data their meaning. This includes those at the highest levels of government as well as those in other roles that affect educational policy.

Advocacy for reauthorization of NCLB, perhaps with a few changes making it ostensibly better, appears to be widespread and bipartisan. The *Washington Post* (2007a) reported support for NCLB from both Representative George Miller and Senator Edward Kennedy and found encouraging the suggestion that state standards be toughened. Another editorial ("Save School Standards," 2007) offered similar support for not "letting schools off the hook" in attempts to "water down" NCLB. Unfortunately, suggestions that NCLB holds schools accountable in a rational way and that state standards should be toughened require either (a) misrepresenting reality or (b) living with a higher rate of failure. Statistical realities may be ignored by members of Congress or anyone else, but that does not make these realities disappear. It only ensures the collapse of the policy based on false assumptions and disaster for those caught in the wreckage, which in the case of NCLB are the children and educators crushed by the ruins.

Implications for Policy Makers and Practitioners

As Gallagher (2007) has noted, educational policy casts long shadows—it influences services, personnel preparation, and research. Policy is not a trivial matter for the practitioner. It affects what teachers are expected to do and what and how they teach, and a given policy is likely to affect educational practice for a generation or more. Thus, it is incumbent upon policy makers and those who advise them to acknowledge all realities, as ignoring any given reality will neutralize policy initiatives, no matter how well intended and no matter whether grounded in reality in most respects.

NCLB is an example of a well-intentioned policy that ignores realities about groups and gaps in educational performance, and it is therefore headed for eventual ignominy (Johns, 2003; Kauffman, 2004, 2005). It ignores the many realities involving distributions of test scores, the reasons for success and failure, and the nature of the differences among groups. The fact that the law was passed with overwhelming and bipartisan support does not make it rational or defensible. Votes may indicate the political palatability of a policy, but they do not make a policy that ignores realities viable. That is, votes do not change all of the realities of the world, only the political reality that they represent.

The implications of this discussion for policy makers and practitioners are simple and straightforward:

- Before you enact or support a policy, make sure that it is firmly and thoroughly grounded in realities about the world of education.
- Resist all attempts, for whatever reasons, to enact or support a policy that you know skirts any reality.
- Do not succumb to the temptation to engage in rhetoric that does not conform to every known reality about education.
- Demand of advisors the kind of data and thinking that will convince reasonable skeptics of the policy's workability.
- Demand consideration of all statistical information, not merely those indices that reflect central tendency.

Some practitioners do not acknowledge realities, partly because their teachers or their policy makers (or both) do not. They have been convinced that realities can be ignored without disastrous consequences. Most practitioners do know fantasy from reality about education. Nevertheless, not all are equipped to articulate all of their quibbles with NCLB. Moreover, many practitioners acquiesce in efforts to implement policies that they know are unworkable because they believe that refusing to go along with pretense will cost them their employment. It is much more difficult to guess why many of those whose employment is not threatened by dissent also acquiesce in the charade of NCLB.

We are not under the illusion that policy must be shown to be flawless before it can be supported. However, supporting a policy *known* to be conceptually faulty is in our opinion unethical. Finding flaws not known to exist at the time a policy is enacted and fixing them are inevitable aspects of policy development. Gillon (2000) pointed out how policies often have unintended consequences, and we do not believe that people can be held responsible for the unpredictable effects of their behavior. Nevertheless, it is also the case that in our society when a person knows (or should know) the negative consequences of an action but performs that action anyway, we consider that person negligent or malicious. Our society does not countenance constructing and marketing goods that are known to be defective. Neither do we fail to hold people responsible for negative outcomes when those people knew or should have known that their actions would harm others. For example, we accept the fact that it is malicious to manufacture and sell vehicles that people know or should have known are unsafe. Our courts generally excuse people from responsibility for the unknowable, but not from the known or knowable. Neither our colleagues in education nor those in politics are free of responsibility when they embrace policies with known deficiencies. We find it incomprehensible that people of average or better intelligence produce or defend education policies like NCLB that so clearly ignore important realities. In most civil endeavors other than education, criminal penalties are exacted or people are found insane for pretending that fundamental realities do not exist. The education of our children is too important to be approached with sloppy thinking, and the consequences of bad thinking about education should be no less serious than the consequences of bad thinking about planes, trains, automobiles, and other things that are important to us.

Conclusion

The improvement of education demands the acknowledgement of basic realities. These realities are things that happen regardless of anyone's wishes or statements; they are features of the world that humans may have discovered and described (or "constructed" through their language), but they will not change because of philosophical preferences or denials. Any philosophy, pronouncement, or policy based on the pretense that these realities do not exist or will not be encountered is doomed to eventual failure, regardless of its social acceptability, emotional appeal, or apparent initial success.

Most people in our society seem to understand that any project in the material

world launched with even a single flawed assumption regarding physical realities will come to naught and may result in very negative consequences. Educators and other people in our society must understand that something similar occurs in the world of education—that any educational project, reform, or policy launched with a flawed assumption regarding reality or a side-stepping of reality is headed for inevitable failure. Rhetoric about education that does not account for basic realities must be labeled the siren call that it is and be resoundingly rejected. NCLB need not be honored. In fact, it should be recognized as another example of silly talk about education.

Written or spoken proposals that ignore realities should prompt imitation of the statement of the late Senator Daniel Patrick Moynihan regarding proposed changes to the Aid For Dependent Children program: "Are there no serious persons in the administration who can say, 'Stop, stop right now? No, we won't have this?' " (Moynihan, 1995, p. A31). Of many reforms and proposals, more people need to say, "Stop! Stop right now! We won't have this!" Frustration with the powers that be is not a sufficient excuse to treat seriously ideas that are not fully grounded in reality (Kauffman, 2002, 2005). Neither is the suggestion that a proposal is known to be partly or mostly right and therefore worthy of support. Only one assumption not grounded in reality is sufficient to render a project or proposal inoperable or disastrous. NCLB is no exception.

Note

Portions of this chapter were previously published by the authors as an article in the journal *Exceptionality*, Vol. 15, No. 2, pp. 75–96 (also accessible at www.informaworld.com).

References

A vote for "No Child": Welcome support in the House for the law that brought accountability to public education [Editorial] (2007). *Washington Post*, p. A12.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Brantlinger, E. A. (Ed.) (2006). *Who benefits from special education? Remediating (fixing) other people's children*. Mahwah, NJ: Lawrence Erlbaum Associates.

Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007). Children left behind in AYP and non-AYP schools: Using student progress and the distribution of student gains to validate AYP. *Educational Measurement: Issues and Practice*, *26*(3), 21–32.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher*, *19*(1), 3–13.

Engelmann, S. (1997). Theory of mastery and acceleration. In J. W. Lloyd, E. J. Kameenui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 177–195). Mahwah, NJ: Lawrence Erlbaum Associates.

Fitzpatrick, A. R. (2007). NCME members write Congress on NCLB. *National Council on Measurement in Education*, *15*(3), 5–7.

Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher*, *36*, 268–278.

Gallagher, D. J. (Ed.) (2004). *Challenging orthodoxy in special education: Dissenting voices*. Denver, CO: Love.

Gallagher, D. J. (2006). If not absolute objectivity, then what? A reply to Kauffman and Sasso. *Exceptionality*, *14*, 91–107.

Gallagher, J. J. (2007). *Driving change in special education*. Baltimore: Paul H. Brookes.

Gartner, A., & Lipsky, D. K. (1989). *The yoke of special education: How to break it.* Rochester, NY: National Center on Education and the Economy.

Gentry, M. (2006). No Child Left Behind: Neglecting excellence. *Roeper Review*, *29*, 24–27.

Gillon, S. M. (2000). *That's not what we meant to do: Reform and its unintended consequences in twentieth-century America.* New York: Norton.

Gladwell, M. (1997, May 19). The sports taboo: Why Blacks are like boys and Whites are like girls. *The New Yorker*, 50–55.

Goodkin, S. (2005, December 27). Leave no gifted child behind. *Washington Post*, p. A25.

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin.* New York: Three Rivers Press.

Heshusius, L. (2004a). From creative discontent toward epistemological freedom in special education: Reflections on a 25-year journey. In D. J. Gallagher (Ed.), *Challenging orthodoxy in special education: Dissenting voices* (pp. 169–230). Denver, CO: Love.

Heshusius, L. (2004b). Special education knowledges: The inevitable struggle with the "self." In D. J. Gallagher (Ed.), *Challenging orthodoxy in special education: Dissenting voices* (pp. 283–309). Denver, CO: Love.

Johns, B. H. (2003). NCLB and IDEA: Never the twain should meet. *Learning Disabilities: A Multidisciplinary Journal*, *12*(3), 89–91.

Kauffman, J. M. (1990, April). *What happens when special education works: The sociopolitical context of special education research in the 1990s.* Invited address, Special Education Special Interest Group, annual meeting of the American Educational Research Association, Boston.

Kauffman, J. M. (2002). *Education deform: Bright people sometimes say stupid things about education.* Lanham, MD: Scarecrow Press.

Kauffman, J. M. (2003). Appearances, stigma, and prevention. *Remedial and Special Education*, *24*, 195–198.

Kauffman, J. M. (2004). The president's commission and the devaluation of special education. *Education and Treatment of Children*, *27*, 307–324.

Kauffman, J. M. (2005). Waving to Ray Charles: Missing the meaning of disability. *Phi Delta Kappan*, *86*, 520–521, 524.

Kauffman, J. M., Conroy, M., Gardner, R., & Oswald, D. (in press). Cultural sensitivity in the application of behavior principles to education. *Education and Treatment of Children.*

Kauffman, J. M., & Hallahan, D. P. (2005). *Special education: What it is and why we need it.* Boston: Allyn & Bacon.

Kauffman, J. M., & Konold, T. R. (2007). Making sense in education: Pretense (including NCLB) and realities in rhetoric and policy about schools and schooling. *Exceptionality*, *15*, 75–96.

Kauffman, J. M., & Sasso, G. M. (2006a). Certainty, doubt, and the reduction of uncertainty: A rejoinder. *Exceptionality*, *14*, 109–120.

Kauffman, J. M., & Sasso, G. M. (2006b). Toward ending cultural and cognitive relativism in special education. *Exceptionality*, *14*, 65–90.

King, L. (2007, June 6). Data suggest states satisfy no child law by expecting less of students. *USA Today*. Retrieved June 21, 2007, from: http://www.usatoday.com/news/education/2007-06-06-schools-main_N.htm.

Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards."* Boston, MA: Houghton Mifflin.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools.* Westport, CT: Heinemann.

Kohn, A. (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan*, *82*, 349–357.

Lawrence, M. (2006, April 7). *The impact of no child left behind.* Presidential invited speech to the American Educational Research Association, San Francisco.

Mostert, M. P., Kavale, K. A., & Kauffman, J. M. (Eds.) (2008). *Challenging the refusal of reasoning in special education.* Denver, CO: Love.

Moynihan, D. P. (1995, September 21). "I cannot understand how this could be happening." *The Washington Post*, p. A31.

Popham, J. W. (2004). *America's "failing" schools: How parents and teachers can cope with No Child Left Behind.* New York: Routledge Falmer.

Rothstein, R., Jacobsen, R., & Wilder, T. (2006, November). *"Proficiency for all"—An oxymoron.* Paper presented at a symposium on "Examining America's Commitment to Closing Achievement Gaps: NCLB and its Alternatives," New York: Teachers College, Columbia University.

Sasso, G. M. (2001). The retreat from inquiry and knowledge in special education. *The Journal of Special Education*, *34*, 178–193.

Save School Standards: Congress should resist attempts to water down the No Child Left Behind Law [Editorial] (2007). *Washington Post*, p. A14.

Schafer, W. D., Liu, M., & Wang, H. (2007). Content and grade trends in state assessment and NAEP. *Practical Research and Evaluation*, *12*(9), 1–25.

Shattuck, R. (1999). *Candor & perversion: Literature, education, and the arts.* New York: Norton.

Winerip, M. (2006, April 5). No child left behind? Ask the gifted. *The New York Times*, p. B7.