

Avoidance of Equivalence by Leveling Challenging the Consensus-Driven Curriculum that Defines Students as “Average”

STEPHEN J. FARENGA

City University of New York, Queens College

DANIEL NESS

St. John's University

RICHARD D. SAWYER

Washington State University, Vancouver

ANTONIO'S FAMOUS LINE FROM SHAKESPEARE'S *THE TEMPEST*, “Whereof what's past is prologue, what to come, In yours and my discharge...” (1610/1880, p. 50) is by no means impervious to the enterprise of schooling and education. Not only are fads in education commonplace; old education policies reemerge decade after decade with the possible inclusion of pseudo-events and new marketing terms that serve as the only intimation of “reform” (Good, Clark, & Clark, 1997; Heckman & Montera, 2009; Lucas, 1999). Moreover, they are often self-serving to the interests of educational for-profit and nonprofit businesses and administrators, and not always those of students (McNeil, 2000). The constant repackaging of educational policy to “help” students succeed demonstrates the failure of past and present so-called reforms.

In the marketing profession, *pseudo-events* are events that are either trumped up to an extent that little, if anything, is either factually accurate or valid, or those that have never occurred to begin with (Boorstin, 1987). Pseudo-events have been used by education policy makers and heads of educational industries in conveying and representing the putative concept of “average” students in public discourse and general advertising. Educational companies and policy makers have used propagandistic tools (Pinar, 2012), such as pseudo-events (Johnson, Johnson, Farenga, & Ness, 2005) and manufactured crises (Berliner & Biddle, 1995) as a means of altering public perception about the everyday academic realities of “average” students, unequivocally the largest demographic of any student group in schools at any point in history.

We argue that the ideological framework surrounding the definition and use of the term *average* as it relates to academic ability is distorted. We contend that education policy makers and semi-regulatory organizations, for both economic and political reasons, have engineered primarily two initiatives that they want the public to embrace as a means of solidifying a

consensus-driven curriculum: 1) Social justice and equality is achieved by categorizing (almost all) student academic performance as *average*; and 2) high-stakes examinations and rubrics are valid substitutes for demonstrating equity and broad field knowledge. These pseudo-events are particularly relevant because they enable policy makers to take a path of least resistance when attempting to show their concern for social equality. In turn, they tend to expend little to no effort in grappling with the problems associated with the required social investments—ameliorating poverty, inequality, racism, and sexual discrimination to name a few (Johnson, Johnson, Farenga, & Ness, 2005; Rothstein, 2004). Historically, the high-stakes component in educational testing maintained a rather low and even negligible profile in standardized assessment. Since the inception of the so-called standards movement in the early 1990s, and surely by the enactment of the No Child Left Behind mandate in 2003, teachers and school administrators have been under increasing pressure to allocate more time during the school day for annual high-stakes test preparation. Moreover, since public officials have become increasingly obdurate in their willingness to finance assessments that provide better long-term returns (for example, those that measure higher-order skills, such as analytical thinking, synthesis, and research skills), testing companies are lowering the academic bar by limiting content and testing basic skills so that students are more inclined to pass high-stakes examinations. Rothstein (2004) notes that the achievement differences that show up on tests of basic skills are quite different than those that show up on tests that measure higher order thinking. Therefore, as pressure increases to prepare students for high-stakes tests, the tests themselves have become less reliable indicators of achievement (Campbell, 1976; Madaus, Russell, & Higgins, 2009; Rothstein, 2004). Due, at least in part, to the lack of organizational representation, educational policies have failed to serve the overwhelming majority of their constituents—*average* students.

We begin by identifying the problem with attempting to categorize the attributes associated with the *average* demographic. The promulgation of the “average” label is explained by a consensus-driven educational strategy that we call “equivalence by leveling.” We then examine the role of rubrics and standardized testing and how these assessment protocols adversely affect *average* students. In doing so, we investigate the groups that benefit from these policies, and how these groups potentially cause collateral damage. We close by offering suggestions toward a more dialectic framework that includes dissensus, conversation, and dialectic interaction as prerequisite criteria in the process of curriculum development (Gershon, 2012; Pinar, 2012).

The Problems of Defining “Average”

One’s attempt to define *average* or the *average student* opens up, at best, possibilities of oversimplification and, at worst, situations that alienate and exclude groups of students. First, average does not mean *average intelligence*. The labels *below average*, *average*, or *above average intelligence* that a student receives at school are not found in nature, but rather, emanate from the mental testing movement that began in the early twentieth century (Borland, 2005; Gould, 1996; Winfield, 2007). Second, given the invention of the average intelligence construct, the term *average* is often misused euphemistically. That is, *average* is, as a substitute for *mediocre*, which possesses negative connotations. Although society seems to allude to *average* in euphemistic terms, we argue that these students—the overwhelming majority—represent some

of the most diverse, multifaceted, differentiated, productive, and potentially successful individuals in any given demographic.

Problems arise, however, when the notion of *average students* is applied to a monolithic group. This has been the case with the onset of rubrics and other assessment devices developed by non-profit and for-profit agencies, and policy makers who attempt to redefine *average* with the use of criterion-referenced benchmarks, rather than that of norm-referenced performance (Rothstein, 2004). The negative implications and outcomes of redefining *average* have been largely ignored by a number of members within the academic community, political leaders, public policy specialists, and special interest lobbyists. Given the multiple complexities associated with defining *average students*, we refer to this demographic as the largest group of students within K-12 schools who exhibit complexly diverse cognitive and non-cognitive abilities. This demographic may change depending on what is being assessed and which method of assessment is utilized. Within the last decade, federally funded education initiatives and agencies have subtly redefined *average* as checkmarks within the second or third of four academic categories, namely, scores of level 2 or level 3 within the typical 1 through 4 point rubric. The metric known as the Percentage of Proficient Students (PPS), which has become an increasingly popular statistic under federal law to demonstrate achievement (Ho, 2008) is closely associated with the rubric scores of 2 or 3.

The economic benefit of redefining *average* and moving as many students as possible into this category has been sanctioned by the U.S. Supreme Court, which ruled that public schools are not obligated to provide additional educational benefits to students already “achieving educationally, academically, and socially” (Board of Education of the Hendrick Hudson Central School District v. Rowley, 1982). This ruling suggests that Congress never intended for a free and appropriate public education to go beyond the rudiments of general knowledge. That is, schools are not required to maximize the potential of children who are not identified as a special population.

Equivalence by Leveling

An analysis of educational policy suggests some disturbing trends. Federal agencies have initiated plans to manage the inequities in society through deceptive means; instead of directly addressing the causes of low-achievement, the government’s education policy seeks to test its way out of achievement differences. The problem that arises is that in order to increase the scores of a select group within a system, other members within that system are held constant (Farenga, Ness, Johnson, & Johnson, 2010; Rothstein, 2004). Learning is a dynamic process in which students have the opportunity to advance. To change the dynamic, a constituent must implement three policies. The first policy for changing the dynamic is to hold some groups constant by creating stasis; that is, keeping students under an academic glass ceiling. A second policy is to set an arbitrary standard and measure growth as a percentage of proficient students—namely, those who meet the standards. A third policy is to limit the curricula. The combination of these three policies creates an outcome and an educational strategy that we call “equivalence by leveling.” The intention of those who practice equivalence by leveling is to create an *average* group so that instructional time is spent on basic education or in remedial mode. This remedial strategy takes precedent over all other concerns and is accomplished with the imprimatur of federal and state governments. The strategy has the greatest negative impact on average students

who could advance, but must remain in a static environment so that their peers can attempt to reach “so-called” average experiential and academic levels. The collaborations that exist among regulatory and semi-regulatory educational agencies advance an agenda that substitutes standards for curriculum and intellectual ability.

Through the use of standards, national and state education enterprises, in complicity with professional associations and testing companies, have thwarted the progress of *average* students. Local and state governments have outsourced their responsibility to an alphabet soup of organizations in order to establish a national stranglehold on what constitutes an appropriate course of study and educational achievement. What should be of concern to state and federal governments is that none of the new standards are research or empirically based (Johnson, Johnson, Farenga, & Ness, 2005). Instead, students are being subjected to curricula that are consensus driven. The vast majority of the public has no idea that a limited number of people determine what is appropriate study for the masses. Mindful questions that have been ignored should have been directed to all educational agencies, requiring them to empirically demonstrate how the new standards are any better than past efforts.

Within the last three decades, the standards movement has been a boon for many educational enterprises at the expense of *average* students and citizens in general. To illustrate, Koretz, Madaus, Haertel, and Beaton (1992) demonstrated the extremely costly endeavor of developing and assessing standards and that, even under limited implementation, this undertaking was estimated to cost \$3 billion annually. Further, Natriello (1998) estimated the costs of the New York City standards initiative to exceed \$3.5 billion, most of which would be allocated to private education agencies. Ravitch (2013) corroborates these patterns and has identified that the city of Los Angeles has committed \$1 billion on technology, such as iPads, in order to implement the assessment of the Common Core State Standards. It should be noted that this was not additional funding from government sources but instead was funding that had been rerouted from a bond issue for the repair of school infrastructure. These examples are three of many that demonstrate the enormous cost involved in initiating new standards.

Due to major cutbacks in numerous state departments of education, a lack of oversight allows educational enterprises or non-governmental agencies to usurp control of curriculum, instruction, and assessment, and to establish new norms for students. These enterprises are either autonomous non-profit or for-profit corporations that parallel the functions of traditional non-profit organizations responsible for education. Moreover, their task is to develop, promote, and enforce standards for students in school systems. Our apprehension is that none of these enterprises are official government entities. Further, when questions arise with regard to construct validity of their assessments, a number of these organizations claim that the information requested is proprietary (NRC, 2001). In a parallel argument, Hittelman (2004) raises concerns about the use of professional organization standards as an outcome measure of educational quality. The upshot, then, is that these constraints make it virtually impossible to substantiate claims made by these organizations. One can easily recognize these educational enterprises through their national brand or the myriad of acronyms used to refer to them. Examples include teacher union related organizations, such as the National Education Association (NEA) and the American Federation of Teachers (AFT); accreditation enterprises such as National Council for the Accreditation of Teacher Education (NCATE) and the Teacher Education Accreditation Council (TEAC); intergovernmental enterprises such as the Council of Chief State School Officers (CCSSO), Interstate New Teacher and Assessment Support Consortium (INTASC), and National Board for Professional Teaching Standards (NBPTS);

educational testing enterprises like Educational Testing Service (ETS), and the College Entrance Examination Board (CEEB); specialized professional association (SPA) enterprises such as the International Reading Association (IRA), National Council of Teachers of Mathematics (NCTM), National Science Teachers Association (NSTA), Association for Childhood Education International (ACEI), and the National Association for the Education of Young Children (NAEYC); and for-profits, which include Pearson, McGraw-Hill, and Harcourt Educational Measurement.

Ironically, a large cross-section of the memberships of these organizations include members of the academy, many of whom have collaborated to gain control of education and establish an audit culture paradigm—a goal that in no way benefits the *average* student and has, in turn, weakened the role of the academy. The tools that they have used to accomplish this feat are standards. Each organization has developed a series of standards that attempt to establish what they believe to be “appropriate levels” of education. While standards have been a staple component of education since its inception in the 1980s, the recent launch of the Common Core State Standards has issued a new chapter in the standards movement—as if students have already mastered the previous standards developed by non-regulatory organizations. Due to budget cuts, non-regulatory organizations are often contracted by state departments of education who outsource their internal workforce (Chingos, 2012; Gilbertson, 2014; Singer, 2015; Weiss, 2015). These organizations now perform regulatory functions by controlling curriculum in education, thus usurping the authority of traditional educational providers, such as local communities, colleges, universities, and states. They achieve their regulatory roles through the standards-based reform and accountability movement. Ironically, this movement began as a result of purportedly failing students in primary and secondary schools (Boyer, 1983; NCEE, 1983; Sizer, 1984), and is now a major cause of limiting student potential by creating the apparition of average achievement.

The Emergent Influence of External Agencies

Studies in organizational psychology demonstrate that organizations exist, first, for their own benefit and, second, to attempt to carry out what is identified in their mission statement (Johns, 1999). This is also the case with different enterprises in education, which include, but are not limited to, test development companies, publishers, specialized professional agencies (SPAs), teacher unions, and political parties. The emergence of for-profit and certain non-profit educational companies has influenced society’s expectations of educational outcomes. We argue that a number of for-profit and non-profit organizations are fervent to address federal educational mandates for self-interest by participating in lucrative funding opportunities offered by federal, state, and private agencies. The purpose of these mandates is to establish a new sense of *average* ability, substantiated by the use of standards and rubrics. This new assessment scheme has lured *average* students into complacency. This subsequently confines and restricts student academic potential as educational and political policies replace the objectives of mainstream education with those of basic or remedial education (Farenga, Ness, Johnson, & Johnson, 2010; Rothstein, 2004). The collateral damage is that *average* students remain in an educational state of uncertainty while teachers’ energies are spent on test preparation for raising below average scores to average scores. In support, Ho (2008) states, “A recent accumulation of quantitative and qualitative evidence supports the hypothesis that ‘bubble kids’—students just below the

proficiency cut score—receive disproportionate classroom attention and make larger score gains under NCLB” (p. 357). One can extrapolate from this finding that due to a lack of attention placed on rigorous study, a large group of students may not be prepared for the rigors of high school and college in spite of their so-called level 3 designation: “meeting the standards.” While each state has its own assessment vocabulary and criteria, the purpose is similar from state to state: to homogenize and categorize academic performance. Further, a recent examination of college entrance examination scores on both the American College Testing (ACT) and Scholastic Assessment Test (SAT) programs demonstrate that less than 45 percent of the students who take these exams are college ready (College Board, 2012; Webley, 2012).

Rubrics: The Tool of Submission

The contemporary standards-based curriculum relies on rubrics for assessment, scoring guides that attempt, but often fail to operationalize a set of standards or objectives. Rubrics are destructive for the average students’ education for a number of reasons. Criticisms of rubrics in the education research literature are ubiquitous (Andrade, 2005; Delandshere & Petrosky, 1998, 1999, 2002; Hillocks, 1997; Johnson, Johnson, Farenga, & Ness 2005; Koretz, 2009; Mabry, 1999; Moskal & Leydens, 2000). To begin with, rubrics either underestimate or overestimate the average student’s ability, thus providing both the parent and the child with an inaccurate and distorted evaluation. Accordingly, rubrics often fail to reflect the appropriate skill level. Second, much of the language that is used in rubric design is fuzzy and lacks clarity of definition. Third, rubrics are often used to quantify either complex behaviors or dispositions—characteristics of academic subjects that simply cannot be evaluated by a single number or term (Delandshere & Petrosky, 2002; Mabry, 1999). Koretz (2009) argues that separating students into categories, such as “below basic,” “basic,” “proficient,” and “advanced” is a potentially misleading measurement and is “one of the worst decisions we made in testing in decades” (p. 2). We suggest that current assessment practices that are supported by state departments of education avoid transparency by using rubrics in order to obfuscate the evaluation process.

Rubric assessment is oppressive, and limits the future freedom of opportunity for the average student (Ho, 2008). Further, many rubrics that are designed to augment the assessment process supply little, if any, additional data to teachers, parents, and policy makers. In reality, many of these rubrics are nothing more than semantic puzzles. Words such as *proficient*, *satisfactory*, *competent*, *sufficient*, *good*, *adequate*, and the like—terms with entirely different meanings—have all been used to describe *average* students who are frequently led to believe that their current level of effort and persistence will be sufficient in future endeavors. However, evidence suggests that this perception of competence is not the case. An examination of college programs suggests that higher education institutions nationwide have a proliferation of remedial courses in reading, writing, and mathematics. Upon the examination of student transcripts, many students average a B grade and pass all of their state assessments. According to data from the College Board, a majority of incoming freshmen are ill-prepared for college and, moreover, their academic gains are limited as their years progress (College Board, 2012; Webley, 2012). Further, Arum and Roksa (2011) conclude that after three semesters of college, students barely show noticeable gains in critical thinking, complex reasoning, and written expression.

States' Sliding Proficiency Scale

One can draw parallels between the methods used for assessing students at the federal level and those used for assessing students at the state level. Finn and Petrilli (2008) discuss the proficiency illusion in which they report that the notion of proficiency is a moving target that varies by state and grade level. In a populous Midwestern state, a student who passed the fourth grade math exam received the level of “proficient,” a term that translates as *average*. However, the cut-off point in this state is so low that this very student scored 83 percent below other students compared to cut-off points in other states. More surprisingly, because the eighth-grade test has more challenging problems and a higher cut-off point to reach proficiency, the student will probably not pass the eighth-grade mathematics test at the same proficiency level when compared to that which was achieved at a lower grade level (Finn & Petrilli, 2008).

In yet another situation, states establish sophisticated sliding-scale scored values to erroneously represent achievement. In a populous eastern state, a student's scale score and performance level on the fourth-grade English language arts test can range from a low of 430 (level 1) to a high of 775 (level 4). Within this range are cut-off points that estimate a student's level of mastery of the learning standards. Average students' scores range from a low of 671 to a high of 721; a level 3 designation. Within the range of these scale scores, any cognitive distinction between students who obtain a scale score of 721 and 722 is impractical. A score of 722 places the student in level 4: “Meeting the Learning Standards with Distinction.” This indicates that: “Student performance demonstrates a thorough understanding of the English Language Arts knowledge and skills expected at this grade level” (New York State Education Department Information and Reporting Services, 2014). The “average” student who receives only one point less (i.e., 721) is placed in level 3: “Meeting Learning Standards.” This indicates that: “Student performance demonstrates an understanding of the English Language Arts knowledge and skills expected at this grade level” (New York State Education Department Information and Reporting Services, 2014). While we challenge the use of rubrics in measuring instruction and assessment, we have a greater concern for the student who obtains a score of 671, barely a level 3, but is nonetheless marginalized due to obtaining a score at a level of so-called accepted competence. Moreover, a student with this score is not entitled to any additional assistance to reinforce his or her skill level.

Similar to the National Assessment of Educational Progress (NAEP), the state examinations cover a restricted amount of content for subjects such as mathematics, language arts, science, and social studies (Johnson & Johnson, 2006). These assessments are not comprehensive samplings of student behaviors from a full course of study, but rather a very limited and selected sample of content. After the first few years, teachers learn what is on the exams and in many cases can use past tests to coach students for upcoming exams or, more passively, fail to correct student cheating (Correa, 2011; Mansell, 2007; Starnes, 2011). Students are then taught format, content, and timing, and are overly conditioned on how to answer a narrow range of problems. The following year, the scores generally increase. We suggest that these students probably learn how to take a specific test and how to determine the number of hours per day to drill in language arts and mathematics at the expense of science, social studies, music, art, and physical education (Farenga, Ness, Johnson, & Johnson, 2010; Johnson & Johnson, 2006; Johnson, Johnson, Farenga, & Ness, 2008).

The reason why many of these state assessments are unreliable is that they frequently fall prey to corruption. Under the proficiency model for student achievement, federal law requires schools to show sufficient annual progress in mathematics and reading by 2014 or face severe penalties (Ho, 2008). A body of research in the field of testing, measurement, and evaluation demonstrates that test scores will rise without any special intervention when faced with the pressure of high-stakes tests. Campbell (1976) argues in *Assessing the Impact of Planned Social Change*:

[t]he more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. (p. 49)

The preceding statement is known in the social sciences as “Campbell’s Law.” While Campbell’s Law supports the utilitarian nature of tests as valuable instruments, tests lose their value as an indicator of educational achievement if they usurp teaching and the development of general competence. . On numerous occasions, Campbell’s Law has provided evidence of wide scale cheating on state assessments in Houston, Chicago, Atlanta, and Baltimore (Beckett, 2013; Green, 2011). Unfortunately, “Race to the Top” and similar federal policies will only exacerbate the problem of increasing the pressure of educational systems to corrupt test score data.

Collateral Damage

We have argued in this paper that “equivalence by leveling” is harmful to students who are labeled *average* based on an unreliable and corrupt system. These students, unable to express unique talents, skills, creativity and dialogic ways of learning within static learning environments, bear the burden of an audit culture and economy. In this section, we explore collateral damage in relation to *average* students and as a result of audit culture. As Berliner (2006) states, “high-stakes testing programs in most states [are proving] ineffective in achieving their intended purposes, and causing severe unintended negative effects, as well” (p. 949). Students experience testing fallout not only in actual testing situations, but also as a result of the pressure that high-stakes testing places on school effectiveness, student cultural identity, curriculum and course offerings, and school climate.

While test scores may represent a valid measure of the quality of schooling for policy makers, as represented, for example, by the use of testing in “Race to the Top,” students become the basis for a bleak—and we argue, hostile—school experience (Triplett & Barksdale, 2005; Wheelock, Bebell, & Haney, 2000). The absurd impossibility of all students actually reaching the “top” of norm-referenced tests (which by their nature require an average) and sharing the collective glory of first place is not lost on students. Research on students’ reaction to high-stakes testing indicates that as students progress through school, their anger levels grow from just over 6 percent of fourth graders to almost 20 percent of eighth and tenth graders who are angry about high-stakes testing (Wheelock, Bebell, & Haney, 2000). As a result of taking high-stakes tests, students “can experience extreme nervousness resulting in psychological, emotional, and physical effects” (Triplett & Barksdale, 2005, p. 256), as well as a diminished sense of empowerment (Cattaneo & Chapman, 2010) and a decrease in feelings of social competence at school (Wang, 2009).

On a curricular level, accountability mechanisms tend to narrow the curriculum (Altoff, 2011; Harris, 2012), replacing learning processes with test taking strategies and by creating a classic deficit pedagogy. Curricula that are formed around hypothetical and abstract end points, such as predetermined outcomes emphasize the ends over the means in the classroom (Cho & Trent, 2005; Clandinin & Connelly, 1992) and label students based on the difference between the desired results and actual score—the deficit. The high stakes assessment process all but guarantees a deficit pedagogy, and test scores label students as they frame their classroom experiences. As students experience a developmental need to explore the relationship between who they are and what they study, students find themselves in an environment of shrinking definitions of the meaning and ownership of knowledge.

This curricular process undermines learning in a number of ways. First, with an emphasis on end results, students' cultural identities and sense of self are disengaged in the classroom. In contrast to promoting a lived or embodied curriculum (Aoki, 1987, 1993; Greene, 1991), such an arrangement decontextualizes the learning process, which becomes generated not by a student-centered-and-animated curriculum (Farenga, Ness, Johnson, & Johnson, 2010), but rather by a curriculum with roots in a remote, centralized location (often located far from the classroom). This process narrows the curriculum to defined and static knowledge, which reflects the form reproduced by students on standardized tests. Most importantly, this “backwards planning” dynamic rejects the emergent and lived process of learning—which is rooted in the narrative, culture, and social identity of the student (Aoki, 1987, 1993; Greene, 1991). Undermining the congruence between the culture of the student and that of the school, we argue, is necessary for both student and school success. Noguera (2012), for example, found that graduation rates of minority males were higher in “safe schools where [the minority males] feel as though they can be themselves, where the peer culture reinforces the value of learning, and where character, ethics, and moral development are far more important than rigid discipline policies” (p. 11).

Second, and ironically, high stakes testing bestows a privileged status on academic subjects that are tested. It further marginalizes those that are not, leading to the elimination of non-tested subjects or to their loss of instructional time (Altoff, 2011; Chapman, 2005). Chapman (2005), for example, reviewing multiple survey data reports that “[i]n elementary schools, test prep and test taking may well exceed the 26 hours typically devoted to once-a-week visual arts instruction in a year” (pp. 132-133). This dynamic limits the opportunity for average students to excel in the truncated or eliminated classes in which they may show exceptional skill and ability. This trend of reducing instruction time in the humanities and the arts was investigated by Klein (2007) who reviewed findings from a study by the Center on Education Policy. Klein states:

In a nationally representative survey of 349 districts, the Washington-based group found that 44percent reported cutting time from other subjects to focus on math and reading. The decreases were relatively substantial, according to the report, totaling about 141 minutes per week across all subjects, or almost 30 minutes per day. (p. 7)

Chapman (2005) describes the dynamic that undermines the arts in elementary schools below:

Although NCLB does include the arts in a list of core academic subjects, the law does little to support education in the arts, or foreign language, or the humanities and social

studies. Indeed, since NCLB has been implemented, these neglected subjects have been called the "the lost curriculum" by the Council of Chief State School Officers (CCSSO, 2002) and cited in a discussion of the "atrophied curriculum" by the Council on Basic Education (2004). (p.118)

As far back as 2004, 25 percent of elementary principals reported cutting arts education and 33 percent anticipated doing so (Chapman, 2005). And since that time the curricular pressure on schools to improve student test scores has only increased. Thus, it is difficult for students who are creative learners or divergent thinkers (Ornstein & Hunkins, 1998) to reverse negative labels of themselves by taking courses such as history, sociology, art, music, filmmaking, or history where their interests and purposes may lie. While this injustice is harmful to these students, it is also harmful to society (as well as the fields of mathematics and science) because it denies imaginative thinking.

In addition to impacting curriculum, new assessment strategies also impact school climate. The change in school climate leads us to our second main concern, which is that the high-stakes component of standardized testing has contributed to the epidemic of bullying found in schools (Goldberg, 2005). In 2009, for example, "28 percent of students age 12-18 reported being bullied at school during the school year" (NCES, 2011) as well as through cyber-bullying (Rigby & Smith, 2011). We contend that these figures on bullying are low, as they do not include high-stakes testing itself as a form of bullying—at least as they affect the lives of students and teachers. Our categorization of the process of high-stakes testing as a form of bullying is based on the impact of testing on students within its concomitant climate of testing, sorting, and labeling. This process meets the three elements of bullying as defined below. Bullying is "repeated exposure to negative action; intention to harm; and imbalance in power" (Olweus, 1993, as cited in Langdon & Preble, 2008, p. 486). Given the loose analogy in this comparison, one might question whether standardized testing is intended to harm students. However, it is impossible to deny that such tests do intend to sort and categorize students. A vivid example can be drawn whenever test scores are publicized and comparisons are made between scores in affluent and poor districts (Farenga, Ness, Johnson, & Johnson, 2010; Rothstein, 2004). We object to the public dissemination of high-stakes test data, arguing that it should remain solely with district and community stakeholders. Researchers can predict the educational outcomes of school districts based on the socioeconomic status of the community alone. We question the purpose of trying to publicly equalize vastly unequal quantities and argue that doing so is a way to bully and humiliate students and teachers from districts with limited resources, and further draws them into compliance.

Analysis of five-digit government issued zip codes is perhaps one of the most evident ways in identifying the stark contrast between low and high socioeconomic status communities, particularly in terms of high-stakes test score correlations. In many communities throughout the country, one can predict achievement based on zip codes alone. This parallels Howard Gardner's (2000) argument that "[w]e can accurately project a child's chances of completing college and her eventual income by knowing only her zip code" (p. 45). Gardner's contention implies an intrinsic relationship between success in high-stakes testing performance, college completion, and income level. Clearly, there exists an imbalance of power that is suggestive of institutionalized bullying.

The argument that school structures might promote bullying is not new. Over half a century ago Jersild (1955) suggested that inequitable student-teacher power dynamics,

exacerbated by a climate of testing and accountability, was the perfect context for teachers to project their own fears, anxieties, and even hostilities onto students. He called for a form of educational iatrogenesis in which teachers and students develop a knowledge of self and others in order to prevent harmful educational routines that lead to bullying. While individual bullying is directly aggressive, institutional bullying is intensely passive aggressive. With rhetoric claiming to help *average* and more disadvantaged students, educational policy characterized by high stakes testing and evaluation, we claim, does the very opposite. These policies create a climate of disadvantage, which becomes a breeding ground for bullying. Granted, it may be difficult to fathom the notion that schools promote disadvantage instead of advantage; and despair instead of hope and possibility. But it is highly plausible that a school culture, expressed through the hostile labeling of students (special, failing, divergent, impoverished, angry, insubordinate, and even perverse and immoral), might turn schools into sites that sanction individual and institutional bullying.

One question that we have examined in terms of bullying is the extent to which *average* students drop out of school because of their perceptions of negative labeling and their accompanying implications (e.g., finding schools hostile). It is important to note that in some cases the term *push out* may be more accurate than dropout. For example, students with low test scores are retained in the ninth-grade to prevent them from taking 10th grade standardized tests (Shriberg & Shriberg, 2006). Students are also suspended or expelled from school (a hostile act) (Fergus & Noguera, 2010; Holmes, 2006), thus impacting lower performing students more than average students. Indeed, the research does suggest a correlation between low test scores and dropout rates (Goldberg, 2005; Platt, 2004; Shriberg & Shriberg, 2006), but not between average test scores and the dropout rate. On the surface, data suggest that the response of *average* students to receiving lower test scores does not result in them leaving school. We need to note, however, that actual high school dropout rates are difficult to determine and have been contested, and also that there is a hidden cultural component (Shriberg & Shriberg, 2006). For example, an uneven distribution of dropout rates exists along cultural lines. Students in specific cultural groups are especially at risk of being pushed out. “[A]pproximately 76.8 percent of Asian students and 74.9 percent of white students finish high school, these figures drop to 53.2 percent for Hispanic students, 51.1 percent for Native American students, and 50.2 percent for black students” (Shriberg & Shriberg, 2006, p. 76). When these data are sorted by gender, their numbers greatly increase for males. Nationally, African-American and Latino males are more likely than any other group to be suspended and expelled from school (Fergus & Noguera, 2010). These figures represent an unambiguous pattern, indicating that cultural factors (exacerbated by poverty) play a role in retention and further suggests that cultural safety issues (arguably for students at all academic levels) exist in schools (Shriberg & Shriberg, 2006). Intensely disturbing in terms of retention alone, data also suggest that many average students in diverse cultural groups, who do not leave school, encounter safety issues.

Since dropping out of school may not represent a viable option for many average students—even those who may feel that they are being exposed to an alienating, desiccated, and punitive curriculum—the question becomes one of, options. Many positive options, such as taking interesting electives or courses like art, drama, music, sociology, and even social studies increasingly do not exist. Instead, students who find their home cultures, interests, talents, and divergent perspectives either ignored, or even labeled as deviant can try to do what unhappy students have done for decades –seek the safe invisibility of being perceived as average. An

intense climate of accountability and labeling, however, is making such an option less tenable. Restricted courses offerings are also making the school environment much less “user friendly.”

What’s Past Is Prologue—Redux

Collateral damage could easily have been predicted by Freire (1970) and Illich (1972), who suggest that the use of bureaucratic standards and criteria may be viewed as a means to which an enterprise gains control of a given situation. Illich posits that schooling devoted to achieving consensus-driven benchmarks (i.e., standards) perpetuates hopelessness for the underprivileged. In our thesis, the underprivileged student populations are those who are misrepresented and underserved. Those who are underserved are a silent heterogeneous majority of students who are ignored by the educational system and thus labeled as “*average* students.” An analysis of the contemporary context places the SPAs and the educational testing agencies at the heart of what Illich would refer to as the most culpable of the bureaucratic enterprises—organizations that serve as gatekeepers and perpetuate stasis. Gate keeping is achieved by requiring the majority of students to conform and adhere to a prescribed conception of average ability. Further extending Illich’s position, the act of labeling students only serves to sectionalize, rather than unite student populations.

An examination of the role of educational gatekeepers demonstrates Freire’s argument in which the “haves” oppress the “have-nots.” Groups who receive recognition by federal or state governments are entitled to extra financial and material benefits, thus placing them among the “haves.” Being among the “haves” enables these groups to establish a social order of dominance whereby they bestow gate keeping authority to members of their organization. This method of funding systematically pits groups and their organizations in opposition, and establishes an entrenchment mentality that has continually caused educational reform to fail.

Further, current assessment practice disenfranchises a large number of students, usually from urban and rural areas, who are academically labeled as *average* and come from less affluent families. Labeling these students as average avoids the need to provide any additional educational services to maximize their academic potential, or to remediate weaknesses that are not identified as severe “enough” to be a learning disability. The average label is simply a matter of economics—a designation used as a cost-savings measure to avoid additional funding by local, state, or federal governments. In this situation, the social construction of average is buttressed by economic concerns, thus establishing a false pretense of acceptable academic performance. The goal is to eliminate the inequities among various groups in society by demonstrating that they are all average. To accomplish this goal, local, state, and federal governments use proficiency-based statistics and the subjective decision when creating cut scores (Holland, 2002). The result, however, sets up another inequity for a large majority of students. In effect, students’ formal education is held at a low level of competence. Groups of educators and policy makers agree to advance an educational climate that embraces low-level performance and mediocrity, and passes it off to an unsuspecting public that putatively accepts it as *average* ability. Vast discrepancies in most state high-stakes test results when compared to federal test results serve as evidence for this behavior (Ho, 2008). Even superficial comparisons reveal how this low level of knowledge that the states consider as *acceptable student performance* develops a false sense of achievement in students (New York State Education Department Information and Reporting Services, 2014). The present educational expectation, then, is based on a minimum competency that is substituted as *average* ability. This practice,

when combined with unequal informal learning opportunities for average low-income students, places them once again at a greater disadvantage than students from more affluent families whose parents are able to provide their children with experiences that make up for the deficient formal education. The research clearly demonstrates that affluent parents have the ability to provide their children with tutors, mentors, educational materials, music lessons, and specialty camps that can both qualitatively and quantitatively increase one's experiences, and which, in turn, can increase academic potential in school (Dearing et al., 2009; Johnson, Johnson, Farenga, & Ness, 2008).

The average student is also a victim of collateral damage. Educational policies that have undergirded the American education system for the last century have consistently countered virtually everything that constructivist researchers and human development specialists have concluded about how children learn. Subsequent cognitive and the education-related research of Lev Vygotsky's work exemplifies how educational policy makers either negate or ignore every principle of research and theory that has established consistent outcomes for enhancing knowledge acquisition and learning (Cole, Cole, & Lightfoot, 2005; Crain, 2005). In fact, Vygotsky himself intended to search for a method that would allow the researcher and practitioner to identify actual (i.e., current) knowledge and potential knowledge (Vygotsky, 1978). Vygotsky's goals and intentions were indeed entirely innovative and forward looking, so much so that his work in cognitive development has influenced researchers in the fields of human development, cognitive development, neuroscience, and education (Bandura, 2006; Cole, Cole, & Lightfoot, 2005; Greenfield, 2000). Thus, contemporary education policy, in particular, its focus on engineering society through testing, ignores past and contemporary cognitive and human development research that informs how students learn.

Conclusion

Broadly speaking, one must recognize that schools do not have one function. In order to be a well-rounded student, both cognitive and non-cognitive skills must be addressed. A better measure of student ability will need to examine the extent to which non-cognitive skills support cognitive skills. Advertisements from prospective employers seek candidates who possess good oral communication skills, collaborative skills, self-discipline, reliability, and persistence. As important as these non-cognitive skills seem, they are ignored and treated superficially by testing programs. The lack of emphasis on these non-cognitive skills seems ironic when considering the high premiums that many employers place on these abilities.

Past and present educational policymakers and administrators have instituted pedagogies that foster a convoluted understanding of the relationship among effort, ability, motivation, and fortuity. These pedagogical perspectives are complicated by parents and caregivers who are often led to believe that their child's persistence and/or effort is sufficient for future achievement and success. Research has indicated that this current method of student assessment, based on a selected level of proficiency "can exaggerate trends, minimize gaps, and more subtly, focus attention on low-achieving students" (Ho, 2008, p. 352).

Taubman (2009) has implicated the standards and assessment movement's negative impact on education through the creation of an audit culture. As a result, the educational philosophies of many schools are modeled after performance goals rather than learning goals. According to Dweck and colleagues (Dweck, 1986, 1990, 2006; Dweck & Leggett, 1988; Elliott

& Dweck, 1988), the ways in which a school models its educational philosophy, either explicitly or implicitly, will have a profound effect on the development of a specific type of learner. Performance centered environments value test scores over knowledge. In contrast to performance goals are learning goals. Environments that advance learning goals teach that competence is the outcome of working through difficult problems in a variety of activities. Effort is also understood as the primary cause of success. We support the tradition of scholarship, which places learning goals within student engagement and experience (e.g., Aoki, Dewey, Dweck, Fitzpatrick, Greene, Noddings, Piaget, and Vygotsky). Important to this tradition is the notion of learning dialectics, which explores opposing ideas and recognizes the value of difference in the learning process (Pinar, 2012). Student safety, imagination, risk-taking, value, identity, and culture are only partial contexts that support student engagement. These schools emphasize a rigorous knowledge-centered curriculum that addresses student failure by teaching students the importance and value of appropriate effort in achieving success (Dweck, 2006). Students who hold this belief are more likely to succeed in high school, and will view success as an ongoing process. For these students, failure is only temporary and they have much to learn through error analysis. Moreover, research indicates that student academic ability (by the end of the eighth grade) is a better predictor of college-level performance than any performance during the four remaining years of secondary school (ACT, 2008). This predictor leads to another factor demonstrating societal culpability with regard to students in the *average* category. In many states, students are led to believe that they have made average progress. They are then left alone for four or so years until the next major assessment. Elementary and middle-school years are critical for the development of concepts and skills for achievement in high school. Consequently, students who barely pass eighth-grade exit exams are at a high risk of not completing a college preparatory program in high school (Finkelstein & Fong, 2008).

Although care is needed when making cross-cultural comparisons, there are lessons to be learned from the Finnish education system. The first lesson is the belief that one should educate the whole child. In addition to learning traditional academic subjects, students in Finland, for example, study cooking and enroll in industrial arts classes. The intention is to provide students with curriculum that encourages a sense of self in society. The second lesson is that the Finnish education system deemphasizes standardized testing as a high-stakes factor throughout a student's educational career. The Finnish system of education does not consider the use of standardized tests for college entrance until the twelfth grade. National examinations are used for research and are not used for ranking schools, teachers, or students. The results of such examinations are used to design support materials for in-service programs that improve educational outcomes (Frederiksson, 2006).

We have discussed how the standards movement, high-stakes testing companies, educational enterprises, and government agencies have established ideological frameworks used to foster the public's putative conception of average achievement. The educational inequities that exist in society have been addressed in numerous publications. Spanning nearly a half century, some of these publications include Johnson and Johnson's (2006) *High Stakes: Children, Testing and Failure in American Schools*, Kozol's *Savage Inequalities* (1992), and *The Coleman Report* (1966). Each identifies the differences that exist in healthcare, nutrition, socio-economic status, guidance, and community resources that combine to negatively impact achievement. Moreover, discoveries in neuroscience demonstrate that deprivation can cause irreversible neurological changes during critical stages in brain development. A number of these changes can cause psychological, behavioral, or neurological problems that hinder educational

achievement. Even though society is aware that these factors can impede development and lower achievement, it does not seem ready to address them in an honest manner, mainly due to their cost. In the short term, to reestablish a norm and create an instrument (a test) that provides immediate empirical evidence of accomplishment is much easier and less costly for policy makers. This way, government officials can state that inequities have been addressed and that the proof is in the test scores. In the long term, however, societal benefits that result from attending to and grappling with the problems associated with poverty (mentioned above) will greatly outweigh the cost.

To close, we refer, yet again, to Antonio's line from Shakespeare's *The Tempest*. We initially used this line to demonstrate how infamous past events are intrinsically associated with current ones, in particular, the inequities of *average* students that are a direct result of inept educational policies. But we would be remiss to ignore the fuller meaning of this line from *The Tempest*. In particular, Antonio's words in context present a broader idea, namely, that we as individuals have a choice; we are not obligated to repeat the past. For our purposes, we can repeat the events of the past by using pseudo-events to deceive the public into believing that all is well when it is really not; or, we can dispense with rubrics and other deceptive measures that policy makers use to segregate students (Johnson, Johnson, Farenga, & Ness, 2005; Moskal & Leydens, 2000). Instead, we can initiate equal opportunities for all students, regardless of their academic levels and in contrast to harmful academic labeling. If the former choice is selected, students and their parents will continue to be misled into thinking that all children, regardless of effort, are treated equally when, in fact, they are being categorized from the day they enter school until the day they graduate. We urge education policy makers to consider the latter possibility as an essential educational alternative.

References

- American College Testing. (2008). *The forgotten middle: Ensuring that all students are on target for college and career readiness before high school*. Iowa City, IA: Author.
- Altoff, P. (2011). The key does not fit all the locks. *Social Education*, 75(3), 130.
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30.
- Aoki, T. T. (1987). Inspiring the curriculum. In W. F. Pinar & R. L. Irwin (Eds.), *Curriculum in a new key: The collected works of Ted T. Aoki* (pp. 357-365). Mahwah, NJ: Lawrence Erlbaum.
- Aoki, T. T. (1993). Legitimizing lived curriculum: Towards a curricular landscape of multiplicity. *Journal of Curriculum and Supervision*, 8, 255-68.
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, 1(2), 164-180.
- Beckett, L. (2013, April 1). America's most outrageous teacher cheating scandals. Propublica. Website: <http://www.propublica.org/article/americas-most-outrageous-teacher-cheating-scandals>
- Berliner, D. (2006). Our impoverished view of educational research. *Teachers College Record*, 108(6), 949-995.

- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Cambridge, MA: Perseus.
- Board of Education of the Hendrick Hudson Center School District v. Rowley, 458 U.S. 176. (1982).
- Boorstin, D. J. (1987). *The image: A guide to pseudo-events in America* (2nd ed.). New York, NY: Vintage.
- Borland, J. H. (2005). Gifted education without gifted children: The case for no conception of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed.) (pp. 1-19). New York, NY: Cambridge University Press.
- Boyer, E. (1983). *High School: A Report on secondary education in America*. New York, NY: Harper & Row.
- Campbell, D. T. (1976). Assessing the impact of planned social change. Paper No. 8 from the *Occasional Paper Series*. Hanover, NH: The Public Affairs Center, Dartmouth College.
- Cattaneo, L. B., & Chapman, A. R. (2010). The process of empowerment: A model for use in research and practice. *American Psychologist* 65(7), 646-659.
- Chapman, L. H. (2005). Status of elementary art education: 1997-2004. *Studies in Art Education*, 46(2), 118-137.
- Chingos, M. M. (2012). *Strength in numbers: State spending on K-12 assessment systems*. Washington, DC: Brookings Institution.
- Cho, J., & Trent, A. (2005). "Backward" curriculum design and assessment: What goes around comes around, or haven't we seen this before? *Taboo*, 9(2), 105-22.
- Clandinin, J. D., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), *Handbook of research in curriculum* (pp. 402-435). New York, NY: Macmillan.
- College Board (2012). *The SAT report on college and career readiness: 2012*. New York, NY: Author.
- Correa, M. (2011). Academic dishonesty in the second language classroom: Instructors' perspectives. *Modern Journal of Language Teaching Methods*, 1(1), 65-79.
- Cole, M., Cole, S. R., & Lightfoot, C. (2005). *The development of children* (5th ed.). New York, NY: Worth Publishers.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Council on Basic Education. (2004). Academic atrophy: The condition of the liberal arts in America's public schools. Retrieved March 17, 2004 from <http://www.c-b-e.org>.
- Council of Chief State School Officers. (2002). *Key state policies on PK-12 education: 2002*. Washington, DC: Author.
- Crain, W. C. (2005). *Theories of human development: Concepts and applications* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Dearing, E., Simpkins, S. D., Bouffard, S. M, Caronongan, P., Wimer, C., Lund, T., Kreider, H., & Weiss, H. (2009) Do neighborhood and home contexts help explain why low-income children miss opportunities to participate in activities outside of school? *Developmental Psychology*, 45(6), 1545-1562.
- Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27(2), 14-24.
- Delandshere, G., & Petrosky, A. R. (1999). Anything can be measured, even colors can be measured: That's not the point. *Educational Researcher*, 28(3), 28-31.

- Delandshere, G., & Petrosky, A. R. (2002). In a contact zone: Incongruities in the assessment of complex performances of English teaching designed for the National Board of Professional Teaching Standards. In C. Dudley-Marling & C. Edlesky (Eds.), *Progressive language practices* (pp. 293-325). Urbana, IL: National Council of Teachers of English.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040-1048.
- Dweck, C. S. (1989). Motivation. In R. Glaser & A. Lesgold (Eds.), *Foundations for a cognitive psychology of education* (pp. 87-136). Hillsdale, NJ: Erlbaum.
- Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Ballantine Books.
- Dweck, C., & Leggett, E. L. (1988). A social cognitive approach to motivation and personality. *Psychological Review*, *95*, 256-273.
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*, 5-12.
- Farenga, S. J., Ness, D., Johnson, B., & Johnson, D. D. (2010). *The importance of average: Playing the game of school to increase success and achievement*. Lanham, MD: Rowman & Littlefield.
- Fergus, E., & Noguera, P. (2010). Doing what it takes to prepare black and Latino males in college. In C. Edley & J. Ruiz (Eds.), *Changing places: How communities will improve the health of boys of color* (pp. 97-139). Berkeley and Los Angeles, CA: University of California Press.
- Finkelstein, N. D., & Fong, A. B. (2008). *Course-taking patterns and preparation for postsecondary education in California's public university systems among minority youth*. (Issues & Answers Report, REL 2008-No. 035). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West.
- Finn, C. E., & Petrilli, M. J. (2008). Conjuring cut scores: How it distorts our picture of student achievement. *American Educator*, *31*(4), 20, 22, 28.
- Frederiksson, P. (2006). *What is so special about education in Finland?: An outsider's view*. Uppsala, Sweden: Institute for Labour Policy Evaluation.
- Freire, P. (1970). *Pedagogy of the oppressed*. New York, NY: Continuum.
- Gardner, H. (2000). Paroxysms of choice. *New York Review of Books*, October 19, 44-49.
- Gershon, W. (2012). Troubling notions of risk: Dissensus, dissonance, and making sense of students and learning. *Critical Studies in Education*, *53*(3), 361-373.
- Gilbertson, A. (2014, August 27). The LA School iPad Scandal: What You Need To Know. National Public Radio. Website: www.npr.org/sections/ed/2014/08/27/343549949/the-l-a-school-ipad-scandal-what-you-need-to-know
- Goldberg, M. (2005). Losing students to high-stakes testing. *Educational Digest*, *70*(7), 10-19.
- Good, T. L., Clark, S. N., & Clark, D. C. (1997). Reform efforts in American schools: Will faddism continue to impede meaningful change? In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *International handbook of teachers and teaching* (pp. 1387-1427). Boston, MA: Kluwer.
- Gould, S. J. (1996). *The mismeasure of man*. New York, NY: W. W. Norton.
- Green, E. (2011, June 23). 2011 school test scores will be clean, Alonso vows. Baltimore Sun. Website: http://articles.baltimoresun.com/2011-06-23/news/bs-md-ci-school-cheating-announcement-20110623_1_abbottston-cheating-school-test-scores

- Greene, M. (1991). Teaching: The question of personal reality. In A. Lieberman & L. Miller (Eds.), *Staff development for education in the 90s: New demands, new realities, new perspectives* (pp. 23–35). New York, NY: Teachers College.
- Greenfield, P. M. (2000). Culture and universals: Integrating social and cognitive development. In L. P. Nucci, G. B. Saxe, & E. Turiel (Eds.), *Culture, thought, and development* (pp. 203-242). Mahwah, NJ: Lawrence Erlbaum.
- Harris, D. (2012). Postscript: Urban schools, accountability, and equity: Insights regarding NCLB and reform. *Education & Urban Society, 44*(2), 203-210.
- Heckman, P. E., & Montera, V. T. (2009). School reform: The flatworm in a flat world: From entropy to renewal through indigenous invention. *Teachers College Record, 111*(5), 1328-1351.
- Hillocks, G. (1997). *How state mandatory assessment simplifies writing instruction in Illinois and Texas*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Hittelman, M. (2004). College by the numbers: Does the new demand for accountability threaten the mission of higher education? *On Campus* (publication of the American Federation of Teachers), December-January, 10-13.
- Ho, A. D. (2008). The problem with “proficiency”: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*(6), 351-360.
- Holland, P. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics, 27*, 3-17.
- Holmes, C. (2006). Low test scores + high retention rates = more dropouts. *Kappa Delta Pi Record, 42*(2), 56-58.
- Illich, I. (1972). *Deschooling society*. New York, NY: Boyars.
- Jersild, A. T. (1955). *When teachers face themselves*. New York, NY: Teachers College.
- Johns, G. (1999). A multi-level theory of self-serving behavior in and by organizations. *Research in Organizational Behavior, 21*, 1-38.
- Johnson, D. D., & Johnson, B. (2006). *High stakes: Poverty, testing, and failure in American schools* (2nd ed.). Lanham, MD: Rowman & Littlefield.
- Johnson, D. D., Johnson, B., Farenga, S. J., & Ness, D. (2005). Trivializing teacher education: The accreditation squeeze. Lanham, MD: Rowman & Littlefield.
- Johnson, D. D., Johnson, B., Farenga, S. J., & Ness, D. (2008). Stop high-stakes testing: An appeal to America’s conscience. Lanham, MD: Rowman & Littlefield.
- Klein, A. (2007). Survey: Subjects trimmed to boost math and reading. *Education Week, 26*(44), 7.
- Koretz, D. M. (2009). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Koretz, D. M., Madaus, G. F., Haertel, E., & Beaton, A. E. (1992). *National educational standards and testing: A response to recommendations of the National Council on Educational Standards and Testing*. Santa Monica, CA: Rand.
- Kozol, J. (1992). *Savage inequalities: Children in America’s schools*. New York, NY: Harper Perennial.
- Langdon, S. W., & Preble, W. (2008). The relationship between levels of perceived respect and bullying in 5th through 12th graders. *Adolescence, 43*(171), 485-503.
- Lucas, C. J. (1999). *Teacher education in America: Reform agendas for the twenty-first century*. New York, NY: St. Martin’s.

- Mabry, L. (1999). Writing the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan*, 80(9), 673-679.
- Madaus, G. F., Russell, M. K., & Higgins, A. (2009). *The paradoxes of high-stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing.
- Mansell, W. (2007). It's not cheating exactly...is it? *Times Educational Supplement*, 4733, 18-19.
- McNeil, L. (2000). *The contradictions of school reform*. New York, NY: Routledge.
- Moskal, B., & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(10), retrieved from <http://pareonline.net/getvn.asp?v=7&n=10>
- National Center for Education Statistics. (2011). Indicators of School and Crime Safety: 2011. http://nces.ed.gov/programs/crimeindicators/crimeindicators2011/ind_11.asp. Downloaded April 21, 2012.
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Research Council (NRC). (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- Natriello, G. J. (1998). *The new Regents High School graduation requirements: Curricular and instructional implications and estimates of resources necessary to meet the new standards*. New York, NY: Community Service Society of New York.
- New York State Education Department Information and Reporting Services. (2014). English Language Arts (ELA) and Mathematics assessment results 2012. Website: <http://www.p12.nysed.gov/irs/ela-math/>
- Noguera, P. A. (2012). Saving black and Latino boys. *Phi Delta Kappan*, 93(5), 8-12.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Boston, MA: Blackwell Publishers.
- Ornstein, A., & Hunkins, F. (1998). *Curriculum: Foundations, principles, and issues*. Toronto, ON: Allyn and Bacon.
- Pinar, W. F. (2012). *What is curriculum theory?* New York, NY: Routledge.
- Platt, R. (2004). Standardized tests: Whose standards are we talking about? *Phi Delta Kappan*, 85(5), 381-387.
- Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Vintage.
- Rigby, K., & Smith, P. (2011). Is school bullying really on the rise? *Social Psychology of Education*, 14(4), 441-455.
- Rothstein, R. (2004). *Social class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. Washington, DC: Economic Policy Institute.
- Sawyer. (2012). Curriculum and international democracy: A vital source of synergy and change. *Journal of Curriculum Theorizing*, 26(1), 22-37.
- Shakespeare, W. (1610/1880). The Tempest. In H. N. Hudson (Ed.). *The complete works of William Shakespeare* (pp. 1-130). Boston, MA: Ginn & Heath.
- Shriberg, D., & Shriberg, A. (2006). High-Stakes testing and dropout rates. *Dissent*, 53(4), 76-80.
- Singer, A. (2015, February 9). Pearson and testing buddies try to shut the barn door. *Huffington Post*. Website: http://www.huffingtonpost.com/alan-singer/pearson-and-testing-buddi_b_6644308.html

- Sizer, T. R. (1984). *Horace's compromise*. Boston, MA: Houghton Mifflin.
- Starnes, B. (2011). Superstars, cheating, and surprises. *Phi Delta Kappan*, 93(1), 70-71.
- Taubman, P. (2009). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*. New York, NY: Routledge.
- Triplett, C., & Barksdale, M. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research*, 37(2), 237-260.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wang, M.-T. (2009). School climate support for behavioral and psychological adjustment: Testing the mediating effect of social competence. *School Psychology Quarterly* 24(4), 240-251.
- Webley, K. (2012). *ACT scores show high school students are not ready for college* (2012, August 24). Retrieved from <http://newsfeed.time.com/2012/08/24/act-scores-show-high-school-students-are-not-ready-for-college>
- Weiss, J. (2015, May 18). Pearson loses most of contract for next four years of STAAR tests. The Dallas Morning News. Website: <http://educationblog.dallasnews.com/2015/05/pearson-loses-most-of-contract-for-next-four-years-of-staar-tests.html/>
- Wheelock, A., Bebell, D., & Haney, W. (2000). What can drawings tell us about high-stakes testing in Massachusetts? *Teachers College Record*, <http://www.tcrecord.org> ID Number: 10634.
- Winfield, A. G. (2007). *Eugenics and education in America: Institutionalized racism and the implications of history, ideology, and memory*. New York, NY: Peter Lang.

