

# Critical Education

Volume 6 Number 1

January 5, 2015

ISSN 1920-4125

---

## *Dominant Discourse, Educational Research and the Hegemony of Test Scores*

Arnold Dodge

*Long Island University*

Ruth Powers Silverberg

*College of Staten Island, City University of New York*

Citation: Dodge, A., & Silverberg, R. P. (2015). Dominant discourse, educational research, and the hegemony of test scores. *Critical Education*, 6(1). Retrieved from <http://ojs.library.ubc.ca/index.php/criticaled/article/view/184561>

### **Abstract**

*Observation of the rapidly changing conversation in school improvement prompted an analysis of public discourse and the creation of a framework for understanding forces both influencing and reinforced by its creation. In this paper, we explicate the interactive processes of normalizing, simplifying and conflating ideas about educational practices, using test score results as a convenient tool. Examination of peer-reviewed articles that focused on “test scores” in the educational research literature (since the inception of the federal No Child Left Behind initiative) through the lens of the discourse provided substantial confirmation of the use of test scores in service of these three processes. The findings of the current study suggest a need for educational researchers (especially those examining structural, instructional, and social programs in schools) to re-examine the purpose and nature of their investigations, not only for the science behind their conclusions, but for the rationale behind framing their investigations in the context of standardized testing. This analysis of the intertwined paths of policy and research may lead to strategies for academics to contribute to the halting of a dangerous and growing trend.*



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and *Critical Education*, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available from <http://creativecommons.org/licenses/by-nc-nd/3.0/>. All other uses must be approved by the author(s) or *Critical Education*. *Critical Education* is published by the Institute for Critical Educational Studies and housed at the University of British Columbia. Articles are indexed by EBSCO Education Research Complete and Directory of Open Access Journals.

## **Introduction**

### *Background/Context*

Since the enactment of the No Child Left Behind law (2002) the modern era of educational reform has focused on accountability – or lack thereof – of the educational establishment. The authors of the current study focus on a particular variety of accountability - the past decade's obsession with test scores. While using research results to inform policy decisions can ordinarily be a dicey proposition, the recent dominance of student test score results as determinants of the success of students, teachers, and public schools puts the educational research community at the center of a high stakes national debate. The authors set out to examine the unchallenged supremacy of test scores in every aspect of the educational discourse and the research that both utilizes and reinforces it. High stakes testing is producing conditions in which all that is taught and learned is dictated by reductive assessments, both inhibiting the goals of public education and nurturing corporate interests, rendering improvement of schooling for all children unlikely. Analysis of the intertwined paths of policy and research may lead to strategies for academics to contribute to the halting of a dangerous and growing trend.

Observation of the rapidly changing conversation in school improvement prompted an analysis of public discourse creation, specifically how the dominant discourse of the last decade is informed by normalizing, simplifying and conflating ideas about educational practices, using test score results as a convenient tool. Examination of peer-reviewed articles that focused on “test scores” in the educational research literature since the inception of the federal No Child Left Behind initiative through the lens of this discourse provided substantial confirmation of the use of test scores in service of these three processes. Quantitative and qualitative tracking of the research on “test scores” also revealed the parallel between political agendas and educational research, illustrating how the research community has served as an unwitting accomplice in creating a constraining climate for education. A critical pragmatism, which has been described as a “rich source for reflection on practice” and as a way to “change us and to make a difference to our practice as researchers and professionals” (Ulrich, 2007, p. 1109), provides the underlying rationale for the work.

A conceptual framework for understanding the context in which the obsession with test scores has evolved emerged from a review of critical analyses of education reform from a variety of sources. The research on “test scores” was then re-examined using the 2002-2011 ERIC document database of peer-reviewed journals. A total of 1,648 studies were reviewed. Studies were categorized in an iterative process concluding in a post hoc review of the variables in the standardized tests sub-group and their imputed relationships. An analysis was conducted on three studies to illustrate how three themes– normalization, simplification and conflation – could also be found as guiding principles in the research.

The authors found that there exists an impressive swath of research territory making claims using test scores as the validation instrument. The conventional wisdom that high test scores mean more learning and better schooling is examined in the context of a public discourse that normalizes behaviors associated with high test scores, simplifies the processes that lead to learning, and conflates test scores with success. The authors make the assertion that researchers should not be in the service of a wider public climate that clamors for easy answers to complex educational matters. The findings of the current study suggest a need for educational researchers (especially those examining testing in schools) to re-examine the purpose and nature of their

investigations, not only for the science behind their conclusions, but for the rationale behind framing their investigations in the context of standardized testing. We further assert that it is possible for researchers to derail the train of standardized testing by refusal to utilize the processes described here.

## Driving Reform

NCLB, a reauthorization of the Elementary and Secondary Education Act (ESEA, 1965), was intended to support disadvantaged students. The NCLB iteration reached into every school in the country, expanding the federal role in education to unprecedented levels. Received by both major parties as potentially transformative – with progressive themes such as disaggregation of test scores along race and socioeconomic lines (Anderson, 2009) – the new law set out to change the landscape of accountability. Student achievement, as measured by results on standardized tests, was at the core of the accountability mandate of the act. Annual testing in reading and math in grades 3-8 was required by 2005-2006 and by 2007-2008 science tests were required at least once in elementary, middle and high school. In addition, there was a requirement that states participate in the National Assessment of Educational Progress (NAEP) in reading and math every other year. The most ambitious goal of all was the requirement that all students be brought up to the proficient level no later than the 2013-2014 school year. The punishment for lack of progress (progress becoming known in NCLB language as AYP or adequate yearly progress) two years in a row included school choice options for parents, private tutoring – at taxpayer expense – and other “corrective” actions, including a mandated change of school governance (No Child Left Behind, 2004). All this has resulted in a pressure cooker environment in which test score results trump all other indicators of success in schools across the country.

The domestic violence perpetrated on schools by the draconian test-based accountability mandates of NCLB has had a parallel track of international competitions for improved scores over the last four decades. In 1967, the International Association for the Evaluation of Educational Achievement (IAEEA) published reports on 12 nations’ high school mathematics achievement. This figure grew to 20 nations by 1978 (Glass, 2008). Similar comparisons followed which today constitute worldwide benchmarking vehicles for test scores. The Trends in International Mathematics and Science Study (TIMSS) first introduced in 1995, the Programme for International Student Achievement (PISA) results first available in 1997 and the Program in International Reading and Literacy Study (PIRLS) installed in 2001 have been used regularly by politicians, the media and the business community to demonstrate just how poorly US schools stack up against other countries. To promote the politically convenient myths that schools are failing and create a charged atmosphere around test score results, the international competitions have proved invaluable (Farhi, 2007). The claims of international differences, often conveyed by scholars, have the effect of convincing policymakers that we have fallen behind in preparing America’s children for the world of work and for higher education (Morgan & Poppe, 2012).

What has been the impact on the public discourse and the research community of a burgeoning emphasis on accountability as measured by test score results, legitimized by a federal education initiative and elevated to a global crisis? What has been the result in terms of priorities in policy, practice and research? The intent of the authors is to instigate reflection on the part of researchers and suggest that we can play a role in stopping the harm caused by high stakes testing to children, teachers, parents, and communities. The current study reviews the

dominant discourse regarding educational accountability, the research on test scores and the cautions involving their nexus.

### **Discourse Matters**

What role does the dominant discourse play in determining research priorities – and vice versa? How does an emphasis on test scores become lodged in the public consciousness as well as in the annals of peer-reviewed journals? How did scholar researchers---who should know better---come to accept test scores as meme? For answers we first looked to the literature on discourse in an attempt to step back from the edge and look at the landscape from a broader perspective. As Guinier and Torres (2002) remind us, it is the atmosphere in the mine that goes unnoticed that kills.

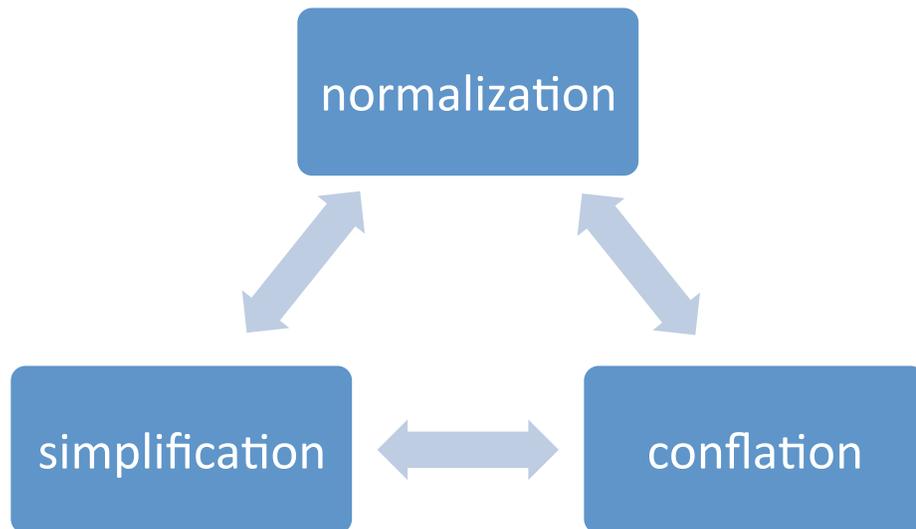
Discourse refers to prevailing language (including symbols, slogans, and images), which frames, focuses attention and renders legitimacy to an issue (Cornbleth, 2008). In education, those who shape the discourse also tend to shape classroom curriculum practice via policy channels. A steady stream of messaging is attended to by educators (and researchers) in their attempt to respond to the “realities” that are a product of the prevailing discourse. A barrage of assaults, remedies and plaintive calls for reform of public schools echoes throughout the culture – in print material, cable TV, the Internet, social media and the like. Cornbleth (2008) reminds us that “...echoes can reduce and simplify, or amplify and embellish...Importantly, they are repeated or reappearing, often becoming a taken-for-granted part of one’s everyday life” (p. 3).

The successful installation of a discourse to explain social phenomena is outsourced through a variety of constructs, which work on the collective mindset of the public. Whether the public weal is served or not is another matter; the prevailing discourse has but one aim and that is to create a climate in which certain types of ideas will flourish and others will be marginalized. The prevailing discourse about state-level standardized testing as accountability...can be seen to delimit, shape, and dominate questions of educational quality, equity, and student achievement. Alternative conceptions of the issues are effectively excluded from the public main stage” (Cornbleth, 2008, p. 9). Kumashiro (2012) asserts that language has the power to “mask certain realities or to di-historicize certain concepts...a range of competing proposals exist on how to reform public schools and yet, in the media, in policy papers, and in speeches by politicians, only certain initiatives seem to count as reform and only certain actors as reformers” (p.10). Whorf (1927), examining the competence of language to convey meaning, suggests that a “... common stock of conceptions...seems to be a necessary concomitant of the communicability of ideas by language” (p. 36). A stock of conceptions, i.e., standards, high stakes testing, choice, competition, ranking, the achievement gap, etc., have made their way into the lexicon of everyday discussions about schooling. This new vocabulary regarding education reform dominates public conversation and stimulates investigative research. The dominance of standards and assessments “crush the system by their sheer weight” (Fullan, 2011, p. 8). Au (2009) avers that the “political legacy” of accountability “operating through high-stakes testing policies, determines the contours and often defines the limits of the political debate. In this regard, any discussion taken seriously within ‘acceptable’ educational policy has to at least recognize, and perhaps cater to, the common sense discourses surrounding test-score productivity” (p. 65).

## Normalizing, Simplifying and Conflating: Discourse and Research Drivers

Researchers, policy makers and the public are caught up in a maelstrom of circular reasoning within the larger discourse context. Researchers find themselves investigating and critiquing practices and approaches to educating youngsters based on test score results, because the discourse is heavily laden with demands for accountability. Armed with research, policy makers are then able to trumpet researched-based initiatives as the answer to educational challenges based on the constructed discourse. The circle is complete. When social science research and educational policy are narrowed and channeled in this way a constriction of ideas and opinions becomes likely, creating a “climate of constraint” which can impede meaningful learning and critical thinking (Cornbleth, 2008). Those who wish to act independently find themselves guided by “social factors that promote or constrain particular expressions of agency” (May and Finch, 2009, p. 538).

The dominance of test scores in the educational conversation creates a constraining climate at the same time that it endorses an orthodox viewpoint. An ensemble of forces – normalizing, simplifying and conflating – found in both the discourse about education and the research examining it is creating a separate reality about schools today (see chart 1). The past decade’s educational terminology, protocols, funding, resource distribution, curriculum planning, etc. are redolent with test-score-driven solutions. These trends are not surprising in light of a public yearning for accountability in their schools and the general belief that accountability for equitable outcomes can be determined with test scores.



*Figure 1.* Constructs Reinforcing the Dominant Discourse Regarding Test Scores

### *Normalizing the Goals*

Normalizing is an integral part of everyday life in both simple and complex cultures. Without normalizing it would be difficult, if not impossible, to function socially and to move the needle of progress. “Normalization processes are found everywhere that people work to implement specific practices and to integrate them in their social worlds. They are important sources of contextual change in organizational settings ...” (May & Finch, 2009, p. 547). Normalizing serves to keep us up to date with our world by constructing “associations that link ideas of circumstances, events, actions, and outcomes that co-occur with some regularity” (Kahneman, 2011).

Normalization, with its implicit valuing, can also be oppressive and constricting. It can become a tool that goes beyond resolving uncertainty and, when manipulated, become a tool for group and individual internalization of “correct” ways to walk, move, talk, act, interact, think, dress, eat, learn, etc. We are found unacceptable or acceptable (sometimes self-normalizing) and look for validation from others regarding our normalcy (Baglieri, Bejoian, Broderick, Connor & Valle, 2011). Ladwig (2010) cautions: “... the more we attempt to gain a better grasp and instrumental control over forms of human life that currently escape the controls of our institutions, our bureaucracies, the more we colonize more of our ‘life world’ within a restrictive logic of systemic input-output prediction” (p. 136).

The power of normalizing is most starkly evident, perhaps, when examining societies and practices that become “normal” in spite of horrific circumstances and dire contexts. Arendt’s “banality of evil” portrait of Eichmann (Arendt, 1963) speaks to great evils in history being perpetrated by the most ordinary of folks (i.e., not the psychopaths and sociopaths) who accepted that what the state asked them to do was “normal.” We normalize the unthinkable quite easily. In a study of evacuation risks, the EPA downplays the “panic image” of human behavior during an emergency situation, insisting that people will stay in threatening situations rather than take a new course of action. Hence, the EPA planning guides for cataclysmic events, e.g., nuclear war, are quite normal in tone (Peattie, 1984).

In the scientific community, normalization is evident in the paradigms that form the assumptions underlying what is investigated and how (Kuhn, 1962). The scientific community assumes that they know what the world is like and those assumptions are defended even at considerable cost, even going so far as to suppress “novelties because they are fundamentally subversive of its basic commitments...Men whose research is based on shared paradigms are committed to the same rules and standards for scientific practice. That commitment and the apparent consensus it produces are the prerequisites for normal science” (p. 11). Aware of the incestuous perils that can lurk behind research, some advise active intervention. In his discussion of pragmatism, Ulrich suggests that a pragmatic research endeavor “...requires us to question the normative content of all of our claims to knowledge and understanding, that is, their unavoidable *selectivity* regarding the ‘practical bearings’ that we consider relevant for judging the object or situation in question” (Ulrich, 2007, p. 1110).

Education is particularly vulnerable to the influence of normalizing. Definitions, goals, regional needs, and financial considerations are diverse and plentiful in education. To bring all this diversity and plenty into focus, narrowing, directing and limiting the outcomes of schooling have become articles of faith for those involved in the conversation. Schools gravitate around normative “centers.” A primary purpose of schooling is socialization to sociocultural norms.

The “normal” child is understood to be the model to emulate, a process that has become so naturalized within schools that its significance is seldom questioned by those who participate in it (Balgieri, et al, 2011). The profile of the normal child has changed over time, but today, “normal” is centered around ways of thinking and knowing and types of knowledge defined by test scores.

Questions about what should be the “norm” in schools is tantalizing fare for education reformers of varying agendas interested in improving curriculum and instruction. The 20<sup>th</sup> century’s growing sophistication regarding measurement techniques supports the normalization of the supremacy of language and mathematics, with the discourse allowing for little regard for other domains, i.e., the arts, second language development, health and recreation. Language and mathematics are tested, and assessment drives instruction. The public perception that academic outcomes are measurable and non-academic outcomes defy measurement feeds the belief in the superiority of literacy and numeracy. The need to demonstrate achievement of normalized/manipulated goals is evident in the amount of time allocated to these subjects. “In this now dominant paradigm, concepts like ‘self-actualization,’ ‘service,’ ‘citizenship,’ and ‘democracy’ are slighted, along with the arts, the humanities, social studies education, and foreign languages” (Byrnes, 2010, p. 2). These vague and difficult to measure constructs are readily and eagerly trumped by test score results in math and literacy which feed into the spectacle mentality – much like sports- where ranking and competition, i.e., winning and losing, fit easily into our culture (Nichols and Berliner, 2008). Again, assessment drives instruction; the appearance of effectiveness is derived from scores on literacy and numeracy tests, and therefore the curriculum must be devoted to that which will produce high scores. A report issued by the Center for Education Policy found that five years after NCLB became law, 62 percent of a representative sample of school districts across the country increased the amount of time (a 47 percent increase in language arts and a 37 percent increase in math) spent on elementary language arts and math. These same districts decreased time allowed for science, social studies, art and music, physical education and recess. In addition, greater proportions of schools identified by NCLB factors as in need of improvement increased time for ELA and/or math than schools not designated in need of improvement. Finally, there was greater emphasis in curricula on state tested content and skills in ELA and math since NLCB inception in schools surveyed (McMurrer, 2007).

The most recent example of a constricting influence of policy based on a new “normal” is the new approach to the evaluation of teacher, principal, and building effectiveness via tests scores sweeping the country. Now that our taken for granted assumption is that “good” and “successful” students are those who do well on tests, we can determine who is good/normal and who is not, and use this distinction to define “successful” and “good” teachers and schools. Findings from the Board on Testing and Assessment of the National Research Council of the National Academy of Sciences (Baker et al., 2010) were unequivocal in their denunciation of the misuse of value added data, test score percentile increases, to measure teachers. Despite the clear and present danger, the test score/evaluation connection continues normalized and unabated. We accept the equivalence of test scores and educational effectiveness and therefore place value on their interrelatedness.

The focus on numeracy and literacy and the standardized testing used to assess them is not only characteristic of US schools but has become the sine qua non of international comparisons (Ladwig, 2010). With the well-disseminated and now popular belief that the United

States underperforms other nations on international tests, American parents have normalized “pressure” on their children. In fact, in a Pew Research survey, 64% of US parents felt that parents were not putting enough pressure on students and 11% felt that there was too much pressure. Results from China were 180 degrees apart from US results, 11% in China saying there was not enough pressure and 68% saying there was too much (Pew Research Center, 2011).

A subtle and particularly virulent strain of normalizing is made manifest in the universally accepted and frequently referenced concept of the achievement gap. Besides contributing a memorable trope to the educational discussion, the achievement gap has become synonymous with the different results found in majority versus minority communities. “Instruments (or, perhaps more accurately, social weaponry) such as aptitude and achievement tests provided (and continue to provide) the blunt force for [an] invasive effort, both visibly and rationally upholding white-superiority ideology over all student populations regardless of race... the constant and continuous comparison of students of color (African and otherwise) to white students as buffered by test scores reinforces those differences in the extreme ...” (Kirkland, 2010, p. 1 ). This academic redlining is made manifest every time a list of test scores is published in the local or regional press. With stunning predictability areas of high achievement are those with white, privileged communities and those with low achievement are typically minority communities with low SES. Boundaries do matter (Wells, 2010) as they reinforce the normalization of success in the wealthier community. Okun (2010) explains normalization in the context of the historical construction of race where white is designated as “civilized, superior, deserving” (p. 6) along with the elevation of the “normal man” (p. 7).

In order to firmly embed into the public discourse (i.e., to normalize) such issues as the supremacy of academic outcomes, a focus on literacy and mathematics, and the presence of an achievement gap, evidence-based investigations provide weighty testimony. May and Finch (2009) suggest that mechanisms to gauge the effectiveness of a new practice are part of the process of normalization: “Regular and organized procedures for monitoring and ongoing assessment of the process and impact of the new practice within an organizational context may involve highly structured and formal mechanisms of institutional knowledge production and interpretation ...” (p. 546). Chomsky (as cited in Ohanian, 2012) goes one step further in describing the politically expedient nature of professional “quarreling” regarding a normalized policy iteration: “The smart way to keep people passive and obedient is to strictly limit the spectrum of acceptable opinion, but allow very lively debate within that spectrum—even encourage the more critical and dissident views. That gives people the sense that there’s freethinking going on, while all the time the presuppositions of the system are being reinforced by the limits put on the range of the debate” (p. 7). For an idea to be firmly normalized, criticism is not only welcome, but becomes integrated into the practice, giving the appearance of healthy self-monitoring: “...both communal and individual appraisal may lead to attempts at reconfiguration in which ideas about the use and utility of a practice are subverted, modified or reconstructed. These play an important part in feeding back into notions of the coherence and meaningfulness of a practice” (May and Finch, 2009, p. 546).

### *Simplifying the Message*

Focus on normalization of test performance provides fertile ground for simplification. The authors assert that the dominant discourse in education over the last decade has a hierarchical nature. For example, in the process of discourse creation, once the educational goals

are normalized, a *simple* message regarding the goals is a powerful engine of reinforcement. By simplifying the metrics which evaluate the goals, the discourse is made more “understandable,” more available to the general public, i.e., more normal.

H.L. Mencken’s dictum that for every complex problem there is a simple, plausible answer that is wrong, appears particularly apropos to the education conversation today. Simplifying complex issues related to teacher effectiveness and student growth is a centerpiece of the prevailing discourse. “Americans are hungry for statistics: easy-to-understand data to explain a very complicated world (Graves, 2002, p. 36)...We have paid so much attention to numerical scores that our appetite craves only more scores ” (p. 38). We look to the numbers, the percentages, and the rankings to answer our most mysterious and ineffable dilemmas. Simple answers have a “cognitive ease” (Kahneman, 2011), which takes little effort to digest. Simple answers become even more powerful if they are readily available. The ready availability of answers provides heuristics, i.e., mental “short cuts.”

Heuristics are woven into the fabric of the discourse on test score results. While useful for explaining complex phenomena, heuristics can lead to severe and systematic errors (Tversky & Kahneman, 1974). A complex phenomenon such as intelligence or achievement expressed in test score results provides an available and convincing heuristic for the unfathomable. The availability heuristic is defined as an oversimplified rule of thumb which occurs when people estimate the probability of an outcome based on how easy it is to imagine as compared with outcomes that are harder to picture and more difficult to understand (Tversky & Kahneman, 1974, as cited in Dodge, 2009). The “availability heuristic” seems especially appropriate to explain the public’s perceptions of standardized testing as a measure of school and student success.

Our predilection for turning the abstract and the mysterious into a unitary entity is emphasized by Gould (1981): “We recognize the importance of mentality in our lives and wish to characterize it, in part so that we can make divisions and distinctions among people that our cultural and political systems dictate. We therefore give the word ‘intelligence’ to this wondrously complex and multifaceted set of human capabilities. The shorthand symbol is then reified and intelligence achieves its dubious status as a unitary thing” (p. 24). Gould goes on to describe “ranking” as the next logical step in a two-step fallacy, i.e., once we have a number associated with a complex construct, we have an obsession with ordering the numbers on a scale for comparison purposes. This same calculus exists as test scores, putatively measuring the complexity of teaching and learning, become rank ordered and showcased in venues from the classroom bulletin board to the international stage.

The dangers of simplifying complex matters seem to go unheeded by those who would promulgate the test-score-as-decider myth. Despite the clear warning from May and Finch (2009): “In real-world studies, predictions about outcomes are complicated by multiple confounders that include the sheer numbers of actors in a process, the weight of numbers and effects of confounding variables, and the intervention of chance ...” (p. 548). For those who would use examples of classes and schools with test score successes for others to emulate miss the point that Willingham (2012) makes: The...“self-evident solution – take what works from one place and implement it elsewhere – is a notorious flop among those who know the history of education policy. Successes depend on many factors that are hard to identify, let alone replicate” (p. 6).

The research community promotes test scores as a meaningful metric when it presents numerical data, which operationalizes complex educational constructs. The multiple choice test, originally conceived of in 1914 as a way to quickly process a national student body of secondary students which had increased dramatically due to the addition of 1.5 million immigrants (Davidson, 2011), remains a staple for sorting and selecting the nation's students. Over the ten-year period since the inception of NCLB the studies that use test score variables to draw conclusions about the success or failure of substantive, nuanced and rich interventions are legion.

### *Conflating the Results*

The final building block of the discourse, one that rests on the normalizing and simplifying components, is conflating the results. While the public may be convinced of what is normal in schools and while they may understand that test scores explain, in simple terms, what students have achieved on examinations, the next important leap in the discourse creation and continuation is to equate test scores with school success in general.

Historically, measurement results have served as a meme, i.e., a cultural item that is transmitted by repetition in a manner analogous to the biological transmission of genes, an idea or element of social behavior passed on through generations in a culture, especially by imitation. Terman's derivative work from Binet's original notions of IQ testing may be the first example of the embedding of a test score number into the public consciousness as an example of cognitive success. Binet warned, "Some recent thinkers... [have affirmed] that an individual's intelligence is a fixed quantity, a quantity that cannot be increased. We must protest and react against this brutal pessimism; we must try to demonstrate that it is founded on nothing" (Engel, 2009, p. 7). The average citizen may be overwhelmed by the nuanced, organic, multi-faceted, and nonlinear nature of a student's educational development. A simpler and more convenient answer fills the void. Politicians, the business community and the media encourage the trade off of complexity for simplicity so that school and student progress can be reduced to "understandable" numbers that appear "legitimate" (Dodge, 2009).

A dominant discourse is sustained if it has the appearance of a legitimizing factor. Rowan (1982) noted that the accountability of schools is fundamentally based upon the extent to which they satisfy the public's perception of legitimacy. Test score results as indicators of school success provide the bona fides to sustain the discourse. Hand in hand with an understanding of the power of the larger framing comes an awareness of what appears to be a doubling down effect on the messaging, an amplifying of already deeply embedded notions in the public consciousness. One way this occurs is the unquestioning relationship between test score results and school success. An almost synonymous relationship exists between the rise and fall of test scores and the weal or woe of school value. This relationship, sadly, exists across all domains of public and professional life.

Beyond individual school success, the stakes have gone international. Perhaps the single greatest and most damaging of the confections existing today is the firmly embedded notion that our economic success as a country is based on our ranking on international tests. Entire public relations campaigns – from corporations, non-profits and even PSAs - have sprung up throughout the media which decry our miserable performance internationally and implore the viewers and readers to do something to shore up our standing. This conflation reaches to the highest offices in the land with the President periodically citing our poor ranking against other countries. The fact that test score ratings are not indicative of economic success or competitiveness or innovation -

these factors determined in much greater measure by levels of outsourcing, tax and poverty rates, health care, natural resources, etc. (Rotberg, 2011) – seems to be irrelevant to those who make the school ranking/economic success connection.

In addition to broad notions of the relationship between test scores and school success, is a troubling sub-category, i.e., the relationship between the test type and the construct validity it is putatively measuring. The sanguine response from all parties when reading scores improve is indicative of a subtle conflation that is at work. Implicit in the reading score improvement is that students have improved in reading when in fact what they have done is improved in taking a reading test. There is no evidence to support that a reading test measures the construct of reading. Will our students enjoy reading more, want to expand their reading horizons, use reading as a tool to understand other disciplines? These questions are rarely asked and, for many, are considered irrelevant. What is clear is the ipso facto regard that educators and the public alike have for test measurements and academic success. Achievement tests have become synonymous with success in all disciplines with a bewildering disregard for construct validity. This was becoming the norm even prior to the formalization of NCLB as pointed out by Smith and Fey (2000): “By now, most public discourse conflates assessment with accountability” (p. 335).

The tabloid press has a particularly toxic strain of messaging regarding test score findings and teacher performance. Two examples from the *New York Post*, a daily with a circulation of over half million, the eighth largest circulation in the country serving a city with the largest school system in the country, demonstrate the message: In an editorial entitled, “Merit Pay Matters,” the paper cites a study which showed that when given a \$4,000 cash incentive at the beginning of the year, with the understanding that a portion of the money would be given back if students’ scores on tests didn’t improve, found that the incentive turned “mediocre” teachers into good teachers and good teachers into great teachers (Merit pay matters, 2012). In a news article entitled, “NYC warned: Top Teachers Getting Away,” a study analyzing attrition in four large urban districts, reported that those districts lose up to 21 percent of their worst teachers and up to 17 percent of their best teachers. The article points out that the teachers were categorized on how well their students performed on standardized tests (Gonen, 2012).

The conflation of great teaching and high scores is a given and goes unchallenged in the tabloid media for the most part, but this distortion resides in the wider environment as well “because the media functions as a window to the outside world...what appears across its landscape actually may become people’s reality” as the message is repeated and echoed (Anastasio, 1999, p. 153). This effect may be all the more intensified regarding the subject of standardized tests: “Standardized tests symbolize the maintenance of order, standards and traditional educational values and practices. The general public also tends to see them as scientific, objective and fair. Echoes that resonate with national narratives and values tend to have considerable staying power” (Airasian, as cited in Cornbleth, 1988). “Tests and the sanctions and rewards that are attached to them convey the public image of fairness, toughness, strong leadership abilities, and the fortunes of office to the policy makers who initiated the program” (Smith and Fey, 2000, p. 342). The narrative plays well for non-educators looking for a multiplicity of simple answers to schooling: the curriculum is aligned and focused on basic skills that are being tested; those students and teachers who are lazy and complacent are incentivized to perform better via test score results; and those who have some hostility to work out towards schools have a convenient cudgel to work with in test score results (Glass, 2008).

Even those today who would argue against test scores, or who have a richer understanding of the influences on teaching and learning, are likely to use test scores in their arguments when promoting a new approach or policy (Kumashiro, 2012). Hargreaves and Fullan (2012) when comparing teachers who had “high social capital” with those who had “low social capital” use gains in mathematics scores to make the point for the importance of social capital; shortly thereafter the authors deride the overuse of testing in the schools. In a strong commentary on the absurdity of the NCLB mandate to produce full proficiency by 2014 - including reference to outside influences other than schooling determining success - Armor (2006) uses disparate test score results as the defining criteria for his argument. In virtually every report of effective schools, reference to achievement based on test score gains – or losses – can be found. No surprise when the dominant discourse requires a bottom line related to hard metrics. Glass (2008) exhorts us to be cautious about the rhetoric of crisis, which gets ginned up by an assortment of spreadsheets and graphs depicting declining achievement scores. It has become commonplace for authors of all genres to allude to this crisis, mainly basing their arguments on the conflation of school success with rising or falling test scores. Peer-reviewed journal authors are no exception.

### **Discourse and Research, a Symbiotic Relationship**

If the discourse on test scores is potent – and ubiquitous – then what indication is there that the research community has been tracking the discourse and providing a corpus of evidence that legitimizes and reinforces it? We argue that change in volume over time of investigations involving test scores is one indicator of this tracking of the lead discourse. Confirmation of the power of the discourse and illumination of more specific elements arose from an exploration of the research catalogued in ERIC under the keyword search “test scores” (limited to only those journal articles which had been subjected to peer review). These were the parameters used to create a context for examining the research community's contribution to the discourse.

An analytical approach was taken to catalogue and categorize publications focused on "test scores" since the inception of No Child Left Behind. Because the Education Resources Information Center (ERIC) is the most comprehensive collection of education research available and because it is the primary source for education researchers, ERIC was the source of our data for mapping the increasing attention to test scores. We set out to examine the volume and nature of articles with test scores as the focus, hypothesizing that there is a circular, reinforcing dynamic between extant public discourse regarding education and the products of research on test scores. In other words, as public demands for accountability increase, research on test scores increases. Evidence drawn from the research continues to fuel the public discourse. As more research is conducted more research is fostered. As more results are produced regarding test scores, more conversations take place.

Investigation of the use of test scores in the research began with a count of peer-reviewed publications fitting the descriptors "test scores" since the passage of No Child Left Behind, the reauthorization of the 1965 Elementary and Secondary Education, the first federal policy to provide funds to schools based on students' SES and requiring accountability for those funds. In order to provide the context for comparison we expanded the count to include the years 1968-2011. A start date of 1968 was chosen because, although the ERIC database came into existence

in 1965, there were no “test score” entries between 1965 and 1968. We discovered a dramatic increase, with the 2001-2011 total at least seven times the 1968-1979 total.<sup>1</sup>

The dramatic rise in “test score” entries since the passage of No Child Left Behind mirrors the escalation of the discourse regarding test scores during the same period. Furthermore, a closer examination of the 1,648 articles published during this time frame shed light on the nature of the research being conducted. While a myriad of topics was investigated over the 10-year span, there were recognizable motifs. Three overarching characterizations emerged: some studies tracked pre-post test scores and the use of criterion referenced tests to measure an educational practice; others offered commentary on the usefulness of test scores, and still others examined independent variables in relation to standardized tests, mostly in math and ELA.

A breakdown of examples of the independent variables used in studies which linked interventions and conditions to standardized tests scores appears in Table 1 below. In this iteration the topics of the studies clustered around three themes: social/motivational, instructional and structural. In these examples, which in some cases can only be described as non-sequiturs, a wide and varied array of interventions and conditions are associated with the results of standardized test scores.

---

<sup>1</sup> The number of entries varied depending on the day/time of the search. This figure is the highest given for the 1968-1978 period for purposes of demonstrating the increase as accurately as possible.

Table 1  
*Examples of Independent Variables Used in Relation to Test Score Results 2002-2011*

Social/Motivational	Instructional	Structural
Advisory programs	Foreign language instruction	Ability grouping
Parent involvement in homework	Piano keyboard instruction	Full day kindergarten
Attitudes, school violence and television viewing	Time in the arts	Pullout program for students with disorderly conduct and potential for math achievement
Attitudes toward competition	Summer academic programs in middle school	Teacher longevity
Transcendental meditation	High school music ensemble presentation	School transfer
Affective response towards social comparison	Physical education	Division of school into themed houses
Aerobic fitness	Computer use at home and in school	8th grade academies
Cash for test scores	1:1 laptop use	Mayoral control of schools
Child care subsidies	Teacher directed student use of technology	Charter schools
Vouchers	Reading and math software products	School board member and superintendent turnover
Interpersonal competence	Change of test format	Class size reduction
Overweight children	Double period of algebra	Military deployment
Conditional cash transfer program	Professional development for high school economics teachers	Public pre-school expenditures
Locus of control	Effect of learning styles	Block scheduling
Match between student and teacher gender	Models of extra-curricular support	Competition between schools
Student absenteeism	Exams read aloud	Partnership with a university

Measuring interventions, conditions or approaches to educational practice against test score results has become commonplace. Often, funding and other resources will be committed to those interventions which yield the best test score results, and conversely, funding and resources will be denied to those interventions that yield poor results. This formula has become a largely unquestioned part of decision making in educational policy and administration. The elevation of test scores as the high water mark for achievement and success in school is an element of public discourse that has been driven into the research arena. The research community responds with effusive gifts of additional reportage.

The accepted definitions of "success" and "achievement" substantially contribute to the conflation of high scores and what have become known as "schools of excellence." The research

implicitly renders "test scores" equivalent to "success;" it logically follows that the higher the scores, the better the school. Here, the use of test scores, with their implicitly normalizing and simplifying influence, serves as a distraction from the real, albeit complex, dialogue about what is a good school. Furthermore, such simplifications provide cover for those who would ignore social and structural issues that underlie the more easily accessible surface dynamics. Herein lies what is perhaps the most troubling and profound dilemma regarding discourse and research on test scores over the last decade. When issues in education involve organic, unpredictable, and immeasurable characteristics, the prevailing discourse and the accompanying research lurch towards simple answers with quantifiable results. This phenomenon is apparent in many of the cases examined for this study.<sup>2</sup>

One such case examined family involvement. The study examined several variables: time spent on homework, student and parent attitudes toward homework, direct family involvement in homework, and achievement outcomes. The introductory section starts - as do many of the articles reviewed - with "achievement" implicitly defined as test scores; the authors don't specifically refer to test scores until the end of the article. This implicit conflation between achievement and test scores is indicative of a normalizing effect as well, i.e., the very mention of achievement has been normalized to mean test scores. The rationale for the study followed this sequence: (1) homework is an important contributor to "student achievement," (2) some students don't do their homework, (3) students and families report tensions around homework, (4) reducing these tensions will lead to several outcomes, one of them being higher test scores. The multi-faceted, highly unique construct of family involvement is viewed in its relationship to test score results.

Another case examined personal traits and school success, illustrating how test scores serve to define paths to ways of being that will lead to success. Specifically, the investigation set out to determine whether competence and risk were associated with academic achievement in rural fifth graders. Surveys were used to develop "behavioral configurations" including "troubled girls," and "model girls," "tough boys" and "model boys." Model boys and girls were friendly, popular, and participated in sports and clubs. The model girls and boys had higher test scores *and* a positive influence on the test scores of their classmates. The scores were used to validate the use of interventions in students' interpersonal competence. The need to simplify an abstruse construct, i.e., interpersonal competence, led to an operational definition of "interpersonal competence" and its connection to test scores.

A study of troubled students investigated the influence these students had on the test scores of their classmates. The rationale was possible application to placement and grouping of students. Issues of mainstreaming emotionally disturbed students in regular classroom settings and its impact on the performance of other students in the class were examined. First, the study found that troubled students have lower test scores. These students affect the achievement of other students, and the more troubled the student, the greater the negative influence. Analysis of the effects connected to other variables revealed that higher income students' test scores are more affected by their troubled peers than lower income students', and white boys are more

---

<sup>2</sup> It was the present authors' decision to eliminate all authors' names from any articles referred to in the ERIC search that was conducted. We do so with respect for their efforts and with regard for the integrity with which they approached their subjects.

affected than black boys. White girls are least affected by troubled peers, and troubled boys have the greatest negative influence. Troubled boys most greatly reduce the potential of higher income students and white boys. The conclusions of this study are disturbing in their potential influence on segregation of students by race, income, and gender, in the service of higher test scores, also known as school success.

In these three cases, the complex nature of family involvement, interpersonal competence, and being “troubled” (with their inevitable companion issues of dysfunction in families, peer pressures, emotional volatility, and the crushing effects of poverty) are at least, in part, reduced to a formula validated by test scores. From charter schools to computers in the home to aerobic fitness the researchers mine for relationships to test score results. From data on student absenteeism (which on its face seems like a reasonable indicator) to boys with girls’ names (which on its face seems absurd) test score comparisons are made. It should be noted that the researchers of the present study have no quarrel with individual pursuits to find relationships in these studies. It is the aggregate that is troublesome, the seemingly endless stream of factors that are grist for the test score mill. This accumulation of data and the sheer momentum of the work done in the last decade on test score results have created a new norm, one that simplifies the complex and in so doing paves the way for dubious conflation. In addition, investigations into other testing phenomena unrelated to standardized tests as well as the stream of commentary about tests continue to flood the research files. The increased use of test scores as the arbiter of success in schooling practices narrows our vision and truncates the potential for more in-depth conversation and understanding of our challenges. Ironically, this very article adds to the deluge.

### **Politics and Science: A Bad Fit**

When researchers set out to investigate a phenomenon there are always precautions to take, e.g., IRB concerns, influence of fund sources, etc. We suggest an additional caveat, one that should be writ large, in the investigations of future education researchers: Is the purpose of the work to advance the science of educational research or advance the politics of the dominant discourse? Further, are my research questions driven by the normalization, simplification, and conflation so readily available as bases? When ideology takes the lead and evidence plays the dutiful servant, we participate in a damaging charade with long-term consequences for the public good. Additionally, when the volume of work “has fattened horribly” (McClintock, 2007), one might ask why so much attention is being paid to a subject. Furthermore, the research community by its very definition is not allied with the political community. Each has its own set of rules and aspirations, very different from one another. Universities, to be true to their mission, must “claim independence from political and commercial interests” (van Wyk & Higgs, 2007, p. 65). While the research community may be focused on examining validity and reliability issues in test score results, the political community is focused on using test score results to prove a point. Those who would use test scores as a tool for political gain cannot wait for “evidence to accumulate,” they need information on demand regardless of its merits (Smith & Fey, 2000, p. 338). Such trespasses should send off alarm bells throughout the research community.

Beyond the caution of using high stakes testing as a tool to promote the dominant discourse, the genre itself has been open to criticism particularly from those who have studied item response theory. Testing has become more about gaming the system, i.e., the ability to interpret the tricks found in tests and navigate through deceptive language, than about a true measure of knowledge gained in a subject (Maly, 2012). The oft-cited Campbell’s Law fits well

here: “The 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 was intended to monitor...when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways” (Campbell, 1976).

### **Proceed with Caution**

Policy and practice, downstream from the researchers’ tower, are where the rubber meets the road. It is in the schools and the communities where we finally see the results of worthy or misbegotten teaching and learning paradigms. A public discourse demanding accountability, and a bloated test score research agenda in its service, have created conditions on the ground that defy common sense. Many have weighed in on the role of the research community vis-à-vis policy considerations. McClintock (2007) laments that “...educational research accumulates in great, growing bulk, with all manner of contradictory findings...” (p. 2). Some have advocated for research on policy matters that is relevant and timely, particularly in light of the assault on public education, this as a bulwark against think tanks that proffer shoddy research (Fuller, 2012). The irony may be that the demand for simple answers distills even the most robust and carefully constructed research down to a solution that may never have been intended. It appears that for many it is simply unfathomable in an era of intense accountability to relinquish the procrustean belief that high test scores equal success in all school experiences.

The overarching conception of our nation’s schools may well determine the nature of the discourse and the accompanying energetic research agenda regarding test scores. Byrnes (2012) references Maxine Greene’s perspective on the matter: “...Greene [refers to it] as the utilitarian purpose of schooling. In this view, business principles are applied to schools, and economics trumps everything. Students are thought of much more as future workers and consumers than citizens. Schools primarily exist to prepare students for the workforce...[it] emphasizes math and science coursework, competition, and job skills.” (p. 2).

The simplification of measurement and the conflation of results in education should serve as warnings every time a researcher writes up the findings of a study. In an atmosphere ripe for simple answers to complex issues, researchers should be particularly cautious about the presentation of their findings. Even in medical research, the “most highly regarded randomized trials...do not guarantee results in individual cases; in the end medical care is about the health and well-being of individual patients...Even the seemingly most determinant causal relationship (such as the relationship between smoking and lung cancer) is really just a probability. How it works out in any individual case will depend on many factors, including genetics, environment, and just plain luck” (Riehl, 2006, p. 26). Drawing “defensible policy conclusions” [is] “fraught with difficulty and controversy” (p. 35) in complex educational contexts (Braun, 2004).

The research community may be wise to rethink its focus on test scores. We need more than one tool to decipher the complexity and promote meaningful conversation about educational reform.

## References

- Airasian, P. W. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education*, 95, 393-412.
- Anastasio, P. A., Rose, K.C., & Chapman, J. (1999). Can the media create public opinion? A social identity approach. *Current Directions in Psychological Science*, 8(5), 152-255.
- Anderson, G. (2009). *Advocacy leadership: Toward a post-reform agenda in education*. New York: Routledge.
- Arendt, A. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. New York: The Viking Press.
- Armor, D. (2006). Can NCLB close achievement gaps? *Teachers College Record*, Retrieved August 6, 2012 from <http://www.tcrecord.org> ID Number 12667
- Baglieri, S. Bejoian, L. M., Broderick, A. A., Connor, D. J., & Valle, J. (2011). [Re] claiming “Inclusive Education” toward cohesion in educational reform: Disability studies unravels the myth of the normal child. *Teachers College Record*, Retrieved August 6, 2012 from <http://www.tcrecord.org> ID Number 16428
- Baker, E., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., et al. (2010). *Problems with the use of student test scores to evaluate teachers* (Economic Policy Institute Briefing Paper #278).
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis and Archives*, 12(1), 1-43.
- Byrnes, R. (2012). Stop linking school improvement, economic competitiveness, and national greatness. *Teachers College Record*, Retrieved June 19, 2012 from <http://www.tcrecord.org> ID Number 16773
- Campbell, D. (1976). *Assessing the impact of planned social change*. Hanover, NH: The Public Affairs Center, Dartmouth College.
- Cornbleth, C. (2008). Echo effects and curriculum change. *Teachers College Record*, Retrieved June 30, 2010 from <http://www.tcrecord.org> ID Number: 15185
- Davidson, C. (2011, September 23). Standardized tests for everyone? In the internet age, that’s the wrong answer. *The Washington Post*, Retrieved September 26, 2011 from <http://www.washingtonpost.com/opinions/standardized-tests-for-everyone?/>
- Dodge, A. (2009) Heuristics and NCLB standardized tests: A convenient lie. *International Journal of Progressive Education*, 5(2), 6-22.
- Engel, A. (2009). *Seeds of tomorrow: Solutions for improving our children’s education*. Boulder, CO: Paradigm Publishers.
- Farhi, P. (2007). Five myths about U.S. kids outclassed by the rest of the world. *Washington Post*, p. B2.

- Fullan, M. (2011, April). *Choosing the wrong drivers for whole system reform* (Centre for Strategic Education, Seminar Series 204).
- Fuller, E. (2012, summer). The work of faculty advocates. *UCEA Review*, 53(2), 17.
- Glass, G. V. (2008). *Fertilizers, pills and magnetic strips*. Charlotte, North Carolina: Information Age Publishing.
- Gonen, Y. (2012, July 31). NYC warned: Top teachers getting away. *The New York Post*, p. 8.
- Gould, S. J., (1981). *The mismeasure of man*. New York: W.W. Norton and Co.
- Graves, D. H. (2002). *Testing is not teaching: What should count in education*. Portsmouth, NH: Heinemann.
- Guinier, L. & Torres, G. (2002). *The miner's canary: Enlisting race, resisting power, transforming democracy*. Mass: Harvard University Press.
- Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York: Teachers College Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kirkland, D. (2010). "Black skin, white masks": Normalizing whiteness and the trouble with the achievement gap. *Teachers College Record*. Retrieved November 2, 2010, from <http://www.tcrecord.org> ID Number: 16116
- Kuhn, T. (1962) *The structure of scientific revolutions* (3<sup>rd</sup> edition). Chicago: The University of Chicago Press.
- Kumashiro, K. (2012). *Bad teacher: How blaming teachers distorts the bigger picture*. New York: Teachers College Press.
- Ladwig, J.G. (2010). Beyond academic outcomes. *Review of Research in Education*, 34(1), 113-141.
- Maly, D. (2012, August 5). Professor finds validity of standardized testing questionable. The Daily Texan, Retrieved August 15, 2012 from <http://www.dailytexanonline.com/university/2012/08/051>
- May, C. & Finch, T. (2009). Implementing, embedding, and integrating practice: An outline of normalization process theory. *Sociology*, 43(3), 535-554.
- McClintock, R. (2007). Educational research. *Teachers College Record*, Retrieved June 26, 2012 from <http://tcrecord.org> ID Number: 13956
- McMurrer, J. (2007). Choices, changes and challenges: Curriculum and instruction in the NCLB era. Washington, DC: Center on Educational Policy.
- Merit pay matters. (2012, August 2). *The New York Post*, p.22.
- Morgan, S. L., & Taylor Poppe, E. S. (2012). The consequences of international comparisons for public support of K-12 education: Evidence from a national survey experiment. *Educational Researcher*, 41(7), 262-268.
- A Nation at Risk (1983) retrieved August 9, 2012 from [http://datacenter.spps.org/uploads/SOTW\\_A\\_Nation\\_at\\_Risk\\_1983.pdf](http://datacenter.spps.org/uploads/SOTW_A_Nation_at_Risk_1983.pdf)

- Nichols, S. L., & Berliner, D. C. (2008). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, Mass: Harvard Education Press.
- No Child Left Behind Act, PL107-110 (2002) (enacted).
- No Child Left Behind (2004) retrieved September 8, 2012 from <https://www.edweek.org/ew/issues/no-child-left-behind/> (first Published: August 4, 2004, updated September 19, 2011).
- Ohanian, S (2012). Against obedience *Critical Education*, 3(9). Retrieved from <http://ojs.library.ubc.ca/index.php/criticaled/article/view/183270>
- Okun, T. (2010). The emperor has no clothes: *Teaching about race and racism to people who don't want to know*. Charlotte, North Carolina: Information Age Publishing.
- Peattie, (1984). Normalizing the unthinkable, *Bulletin of the Atomic Scientists*, 40(3), pp. 32-36,
- Pew Research Center. (2011, August 23). Americans want more pressure on students, the Chinese want less, The parenting gap: U.S. versus China (Pew Global Attitudes Project). Washington, D.C.
- Riehl, C. (2006). Feeling better: A comparison of medical research and educational research. *Educational Researcher*, 35(24), 231-29.
- Rotberg, I. (2011, September 14). International test scores, irrelevant policies. *Education Week*, p. 17.
- Rowan, B. (1982). Organizational structure and the institutional environment: The case of public schools. *Administrative Science Quarterly*, 27, 259-79.
- Smith, M.L., & Fey, P. (2000). Validity and accountability in high stakes testing. *The Journal of Teacher Education*, 51(5) 334-344.
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Ulrich, W. (2007). Philosophy for professionals: critical pragmatism. *The Journal of Operational Research Society*, 58(8), 1109-1113.
- van Wyk, B. & Higgs, P. (2007). The call for an African university: A crucial reflection. *Higher Education Policy*, 20, 61-71.
- Wells, A. (2009). *Why boundaries matter: A study of five separate and unequal Long Island school districts*. New York: Teachers College, Columbia University.
- Willingham, D.T. (2012). Measured approach or magical elixir: How to tell good science from bad. *American Educator*, 36(3), 4-12, 40.
- Whorf, B. (1927). On the connection of ideas, In John B. Carroll (Ed.), *Language thought and reality: Selected writings of Benjamin Lee Whorf, published 1956* (p. 36). Cambridge, MA: MIT Press.

### **Authors**

Arnold Dodge is Associate Professor and Chair of the Department of Educational Leadership and Administration at LIU-Post.

Ruth Silverberg is Associate Professor of Education at College of Staten Island, CUNY.

# Critical Education

[criticaleducation.org](http://criticaleducation.org)

ISSN 1920-4175

## **Editors**

Stephen Petrina, *University of British Columbia*  
Sandra Mathison, *University of British Columbia*  
E. Wayne Ross, *University of British Columbia*

## **Associate Editors**

Abraham P. DeLeon, *University of Texas at San Antonio*  
Adam Renner, 1970-2010

## **Editorial Collective**

Faith Ann Agostinone, *Aurora University*  
Wayne Au, *University of Washington, Bothell*  
Marc Bousquet, *Emory University*  
Joe Cronin, *Antioch University*  
Antonia Darder, *Loyola Marymount University*  
George Dei, *OISE/University of Toronto*  
Stephen C. Fleury, *Le Moyne College*  
Kent den Heyer, *University of Alberta*  
Nirmala Erevelles, *University of Alabama*  
Michelle Fine, *City University of New York*  
Gustavo Fischman, *Arizona State University*  
Erica Frankenberg, *Penn State University*  
Melissa Freeman, *University of Georgia*  
David Gabbard, *Boise State University*  
Rich Gibson, *San Diego State University*  
Dave Hill, *Anglia Ruskin University*  
Nathalia E. Jaramillo, *University of Auckland*  
Philip E. Kovacs, *University of Alabama, Huntsville*  
Saville Kushner, *University of Auckland*  
Zeus Leonardo, *University of California, Berkeley*  
Pauline Lipman, *University of Illinois, Chicago*  
Lisa Loutzenheiser, *University of British Columbia*  
Marvin Lynn, *University of Illinois, Chicago*  
Linda Mabry, *Washington State University, Vancouver*  
Sheila Macrine, *Montclair State University*  
Perry M. Marker, *Sonoma State University*  
Rebecca Martusewicz, *Eastern Michigan University*  
Peter McLaren, *University of California, Los Angeles*  
Brad J. Porfilio, *Lewis University*  
Stuart R. Poyntz, *Simon Fraser University*  
Kenneth J. Saltman, *DePaul University*  
Özlem Sensoy, *Simon Fraser University*  
Patrick Shannon, *Penn State University*  
Kevin D. Vinson, *University of the West Indies*  
John F. Welsh, *Louisville, KY*