

# Accountability as a Design for Teacher Learning: Sensemaking About Mathematics and Equity in the NCLB Era

Urban Education

1–27

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0042085916646625

uex.sagepub.com



Ilana Seidel Horn<sup>1</sup>

## Abstract

Using a learning design perspective on No Child Left Behind (NCLB), I examine how accountability policy shaped urban educators' instructional sensemaking. Focusing on the role of policy-rooted classifications, I examine conversations from a middle school mathematics teacher team as a “best case” because they worked diligently to comply with the NCLB. Using discourse analysis, I identify instances of *torque* in their conversations: when educators' compliance with accountability logics pulled them away from humanistic goals of education in ways that stood to exacerbate existing educational inequality. This article contributes to documentation on unintended consequences of accountability policies while identifying features that contribute to torque.

## Keywords

middle school, mathematics, No Child Left Behind, case studies, academic achievement, urban education, teacher development

## Introduction

The architects of No Child Left Behind (NCLB; U.S. Congress, 2001) sought to increase student achievement in public schools. Through annual testing of

---

<sup>1</sup>Vanderbilt University, Nashville, TN, USA

### Corresponding Author:

Ilana Seidel Horn, Department of Teaching and Learning, Peabody College, Vanderbilt University, Box 230 GPC, 230 Appleton Place, Nashville, TN 37203, USA.

Email: [ilana.horn@vanderbilt.edu](mailto:ilana.horn@vanderbilt.edu)

all students, the impact of teaching on student learning could be laid bare for appropriate rewards and sanctions. The consequences for not performing satisfactorily—in the jargon of the policy, not “making AYP” (adequate yearly progress)—involve a loss of autonomy via reconstitution, chartering, or state takeover. In this way, NCLB sorts schools into categories that have consequences for educators, students, and communities.

NCLB requires schools to show performance improvements within different student subgroups, demarcated by children’s racial, linguistic, poverty, and special education status. Perhaps not surprisingly, urban schools have been disproportionately affected by NCLB’s harsher classificatory consequences (Darling-Hammond, 2007). Due to demographic shifts, U.S. cities are more segregated than they were 50 years ago (Ware, 2002), resulting in hypersegregation that leaves children of color overrepresented in many urban schools. As testing bias frequently results in distorted reports of learning, the policy’s requirement that children from historically disenfranchised groups do not get “left behind” paradoxically increases the likelihood that their schools will face sanctions (Milner, 2012; Sunderman, Kim, & Orfield, 2005). Meanwhile, economically privileged communities have the means to insulate themselves from these harsher consequences (Malen, 2011, as cited in Trujillo, 2013), exacerbating the very inequality that NCLB seeks to remedy. Given these conditions, it is no wonder that scholars use words like “irony” (Darling-Hammond, 2007), “contradiction” (Gay, 2007), and “paradox” (Beachum, Dentith, McCray, & Boyle, 2008) to describe NCLB’s impact on urban schools, as the policy often intensifies the problems it purports to address.

The purpose of the present analysis is to point to one critical site where NCLB’s paradoxes manifest, where color-blind, techno-rational policy discourse makes contact with teachers’ and students’ lives. In particular, I look at how NCLB classifications shape mathematics teachers’ instructional sensemaking at an urban school whose leaders worked diligently to comply with the policy’s mandates. My goal is not to set up these educators as straw men and set them aflame but rather to examine a school where NCLB’s logics were carried out with fidelity in ways directly linked to problematic instructional decisions. By understanding how a faithful implementation ultimately worked against the purported aim of increasing educational equity, this “best case” highlights the gaps in policy design and implementation, identifying sources of these contradictions.<sup>1</sup>

Classification is a productive site for analyzing NCLB’s impact on teachers’ work because it is a hallmark of bureaucratized systems, a ubiquitous feature of modern life, and a place where institutional logics meet everyday practice and sensemaking. As Geoffrey Bowker and Susan Leigh Star (1999)

argue in their book *Sorting Things Out*, to classify is human. However, classification systems are not neutral, rational tools for simply organizing the complex mess of reality. Instead, they carry values, silences, epistemologies, and, in many cases, material consequences to individual lives. Through modern institutions, classification systems have become so commonplace as to become invisible, often putting them outside of our analytic reach. Yet their potential to collide with biographies and contribute to oppressive systems makes their analysis imperative for those who want to better understand how social systems and individuals interact.

NCLB's classificatory infrastructure has become especially consequential in the lives of mathematics teachers, who work in a frequently tested—and frequently troublesome—school subject. As a result, the policy's contradictions are pervasive in their work, making it a productive place to locate this analysis. By closely examining moments of *classificatory collision*, as mathematics teachers make decisions in an accountability framework, the current mechanism for instructional improvement via NCLB—increasing rates of proficiency on standardized tests—reduces teaching's “endemic uncertainty” (Lortie, 1975) by eliding a fundamental purpose of education: namely, its critical, humanizing goals (Gay, 2007). This shift away from a view of teaching as humanistic work toward a view of teaching as technical work results in frequent distortions—what Bowker and Star call *torque*—by incentivizing less responsive forms of instruction. The consequences of this move may be particularly detrimental to urban students from historically disenfranchised groups, who already bear undue burden in the institution of schooling (Ladson-Billings, 2013). This, in turn, increases the oft-noted ironies within an accountability policy that—in name at least—strives for equitable outcomes. The framework offered here for connecting accountability policy and instructional decision making stands to inform other accountability policies, such as *Race to the Top* and *Every Student Succeeds*.

## **NCLB as a Learning Design for Educators: Policy as Learning Design**

Policies can be seen as an intentional attempt by members of one group to influence the practices of members of another group (Coburn & M.K. Stein, 2006). Analysts increasingly recognize that educational policy implementation involves active sensemaking by teachers and other practitioners, thus implicating their extant understandings of core aspects of education, such as content, teaching, students, and learning (Spillane, Reiser, & Gomez, 2006; M.K. Stein & Nelson, 2003).

Cobb and Jackson (2012) take these observations about policies and sensemaking a step further, proposing a *learning design perspective* on policy. They suggest that, analytically, policies' underlying theories of action can be looked at for how they organize learning related to changes in practice. In particular, they suggest that we can examine the *why*, *what*, and *how* of policies to understand their possible supports for learning. As these adverbs suggest, the *why* concerns the rationale for enabling policy implementers to change, the *what* entails envisioned outcomes, and the *how* addresses supports for achieving them.

Looking at NCLB from a learning design perspective, the *why* is self-evident: The policy's architects sought to remediate longstanding inequities in educational outcomes in the United States. Likewise, the *what* is fairly straightforward: These remediated inequities would result in increases in student achievement.

The *how* of NCLB, as it turns out, is the source of many contradictions and is therefore a productive place to untangle NCLB's snarled web of learning supports. The policy's *how* is, from a learning design perspective, at once overly narrow and alarmingly underspecified. At the one extreme, standardized tests are the sole means for documenting student learning, narrowing not only the instruments for but also the possible narratives of successful teaching. At the other extreme, the means for improving student learning are completely indeterminate. Pragmatically, this instructional agnosticism keeps the federal government out of heated (and often highly political) debates about best teaching practices (see, for example, Schoenfeld, 2004). In the end, however, the highly specified measures coupled with unspecified instructional practices form a powerful set of countervailing forces for educators to negotiate.

### *What the How of NCLB Specifies: Educational Inequality as a Technical Problem*

Because it implicates standardized test performance without direct supports for instruction, NCLB effectively treats unequal educational outcomes as a technical problem (Gutiérrez, 2013). To be sure, some aspects of NCLB's *how* were patently clear: Annual standardized tests would document student learning and, by extension, school performance in remedying inequality. Schools failing to show progress would then face oversight or penalties such as reconstitution. The mandate for annual testing undid the compensatory nature of the earlier *Elementary and Secondary Education Act* (U.S. Congress, 1964), with any potential for changes in instruction remaining largely

unfunded (Darling-Hammond, 2007). Instead of investing in instructional improvement systems, then, the incentive to improve rested on the threat of sanctions for failing to meet performance targets—all stick and no carrot. This design left the most economically vulnerable schools at the greatest mercy of the accountability regime (Diamond & Cooper, 2007).

From an equity standpoint, the reliance on standardized tests in itself opened up a host of problems. As many scholars have noted, NCLB's techno-rational rhetoric masks the underlying reasons for the underperformance of students in urban schools on these measures. First, standardized testing itself has a troubled history in regard to race and class (Au, 2007, 2013). Early on, IQ tests were regularly used to track African American students into vocational education by European American teachers, thus rationalizing students' low academic performance (Stoskopf, 1999). Relatedly, the ostensible objectivity of these tests appeals to a naive sense of meritocracy while ignoring the "educational debt" owed to underserved students in an unequal educational system (Ladson-Billings, 2013). The underlying logic rests on a premise of valid and reliable psychological measurements for learning potential, which, as a consequence, naturalizes structures like tracking as well as results like unequal student outcomes. Although standardized tests differ from IQ tests in that they do not purport to account for intelligence, they do claim to provide "objective" evidence of student learning—despite well-known validity and reliability issues in generalizing across all student populations. For example, validity threats to standardized testing include schools' unequal resources (Darling-Hammond, 2007), linguistic and cultural bias in the instruments themselves (Gándara & Baca, 2008; Jencks & Phillips, 2011), and stereotypes about ability that influence performance (Steele, 1997), all of which are exacerbated in the case of underserved youth. The focus on individuals' performance effectively erases these broader systems at play, supplanting concerns about equity with deracialized discourses about educational achievement (Leonardo, 2007) while reinstating the narrative of educational meritocracy (Au, 2013). Leonardo describes this irony pointedly, explaining that "by ostensibly giving public schools a chance to show progress, NCLB gives whiteness the license to declare students of color failures under a presumed-to-be fair system" (2007, p. 269).

This "declaration of failure" and its consequences are often keenly felt in mathematics classrooms. Standardized measures assume all students start from the same point in their measurements of "growth." This premise is flawed for a number of reasons. To begin with, there is a shortage of strong mathematics teachers in urban schools (Ingersoll & Perda, 2010), an issue that cumulatively limits many children's learning opportunities year after year. Even beyond school staffing challenges, technical issues abound in the

measures themselves. First, proficiency cut points are arbitrary in that they are not linked to validated notions of learning, and they are politically influenced (Ho, 2008). In addition, the idea of having all children become “mathematically proficient” remains a questionable goal if it is based on measurements alone. Using a definition of proficiency linked to the National Assessment of Educational Progress (NAEP), the most constant measure we have of student achievement in the United States, Linn (2003) calculated that, even at the fairly brisk rate of progress made during the 1990s, it would take about 61 years to get “all students proficient” in Grade 8 mathematics and 166 years in Grade 12. Whether viewed from a perspective of unequal resources, testing bias, or technical flaws, the proficiency game is rigged. Linking proficiency targets to standardized test performance only intensifies the gatekeeping role that mathematics plays in the school curriculum. The techno-rational *how* of NCLB, then, is particularly problematic for urban schools striving for educational equity.

### *What the How of NCLB Does Not Specify: Equitable Mathematics Instruction*

NCLB’s *how* leans heavily on standardized test performance without specifying instructional means. Ironically, the specified *how* of standardized testing may in fact drive teachers away from identified equitable practices in mathematics instruction. For instance, teaching approaches shown to increase mathematical access focus on student sensemaking (Boaler & Staples, 2008; Gutstein, 2003; Horn, 2012; Tate, 1995), a practice often at odds with the strict adherence to pacing guides designed to align instruction and testing blueprints. As pacing guides focus teachers on issues of “coverage” over “understanding,” topics get superficial treatment, as educators keep an anxious eye on standardized test performance. As a result, teachers eschew the richer mathematical tasks that support deep sensemaking (Gutiérrez, 2013; Gutstein, 2003; Horn, 2012; Moses & Cobb, 2001; Tate, 1995). In addition, equitable mathematics instruction engages students’ sense of self (Battey, 2013; Boaler & Greeno, 2000; Wilson, Nazemi, & Jackson, in preparation) with an eye toward positioning children as competent (Dunleavy, 2015; Gresalfi, Martin, Hand, & Greeno, 2009; Gutiérrez, 2013; Horn, 2012). The pervasive labeling of children via proficiency categories stands to supplant richer views of mathematical competence with discrete results from standardized test performance. In addition, the color-blind discourse of NCLB masks racial stratification, supporting narratives of mathematics as “culture free.” Yet researchers have documented how students experience mathematics classrooms as racialized

spaces, where some students are positioned as more competent than others as soon as they step through the door (Martin, 2009; McGee & Martin, 2011; Nasir & Shah, 2011), making teachers' deliberate efforts to resist deficit notions of race, ethnicity, language, poverty, and mathematical ability central to equitable mathematics instruction.

This suspected erosion of equitable instructional practices as a response to NCLB has been identified in other studies. In direct opposition to these equitable mathematics pedagogies, NCLB and its threat of sanctions have narrowed curriculum and teaching in many urban schools (Au, 2007; Darling-Hammond, 2007; Ellis, 2008). Loveless (2013) also found a steady increase of tracking in upper elementary school mathematics that may be a response, as NCLB's design serves to "focus educators' attention on students below the threshold for 'proficiency' on state tests" (p. 20) and provide justification for grouping struggling students.

### *NCLB as a Learning Design: Classificatory Infrastructure and Educators' Sensemaking*

Many of the tensions resulting from the *how* in NCLB can be seen in the way its classificatory infrastructure operates in the daily work of schools. The practices around NCLB categories shape educators' work and their sensemaking (Horn, 2007).

Bowker and Star (1999) define a classification system as,

a spatial, temporal, or spatio-temporal segmentation of the world. A "classification system" is a set of boxes (metaphorical or literal) into which things can be put to then do some kind of work—bureaucratic or knowledge production. (p. 10)

The authors analyze a multitude of classification systems, from the International Classification of Diseases (ICD) to the classification of race under South African apartheid. They argue that such categorization practices cannot truly meet the abstract ideal of sorting things out. The world is too messy and full of cases that can only be given classificatory placeholders like "other," and these exceptions can often bear the brunt of a bureaucratic system.

Like other classification systems, NCLB's classifications have become bureaucratized and ubiquitous. They have material and symbolic manifestations, as when schools failing to make AYP goals get reconstituted or chartered. Schools' AYP status comes out of another set of test-based classifications, the sorting of students based on standardized test scores.

These assessments, per the policy, sort children into categories like *growth*, *bubble*, *proficient*, and *commended*, with students in the latter two categories deemed as having met learning targets measured on the tests. The *growth* category indicates that a student has a long way to go before he or she will become *proficient*.

A central thesis of this analysis is that NCLB's materially consequential classificatory scheme has normalized triage in teachers' work even though, from a perspective of humanistic education, these practices are often highly problematic. *Triage* refers to the disproportionate investment of instructional resources squarely on students close to proficiency cut points—the bubble kids—potentially at the expense of other students and their educational needs.

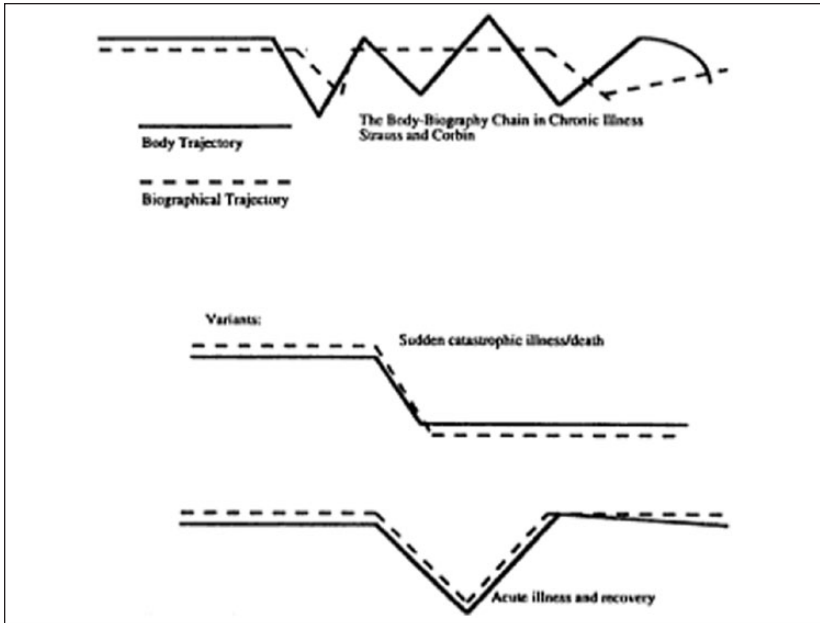
### *Reinstating Work Practice: Educational Triage Versus Humanistic Education*

Why is triage raising concerns among educators? In the framework I propose here, there is an epistemological clash between the classificatory scheme of NCLB and the ideals of humanistic education. By humanistic education, I refer to the stance that education should reach beyond the transmission of knowledge into the development of full human beings. In a democratic society, this form of education aims to develop students' empathic capacity as they learn to weigh alternative perspectives and develop a sense of agency in the world (Bruner, 1961). Teachers, in this view, should foster children's intellectual, moral, and ethical potential while respecting and affirming their cultural identities (Gay, 2010; Ladson-Billings, 1995; Milner, 2010).

The situated nature of humanistic education brings with it a deep uncertainty, having been described as not lending itself to prescriptive instruction (Ladson-Billings, 2011), full of "dilemmas" (Ball, 1993) and "predicaments" (D. K. Cohen, 2011). Humanistic education, then, has an epistemology of ambiguity, which bumps up against the techno-rational positivism of NCLB's classification system and related practices. As I shall explain, these conflicting forces work together and contribute to educational torque.

*Educational torque.* Bowker and Star describe what happens when individual trajectories fall prey to torque, as people's biographies become twisted in response to the distorting forces of classification. To look at the interaction between individuals and classification systems, Bowker and Star expand on Strauss and Corbin's (1997) representations of the relationship between body and biography in illness. The body's trajectory is tied to but is not always completely aligned with the identity trajectory of a patient. They used the



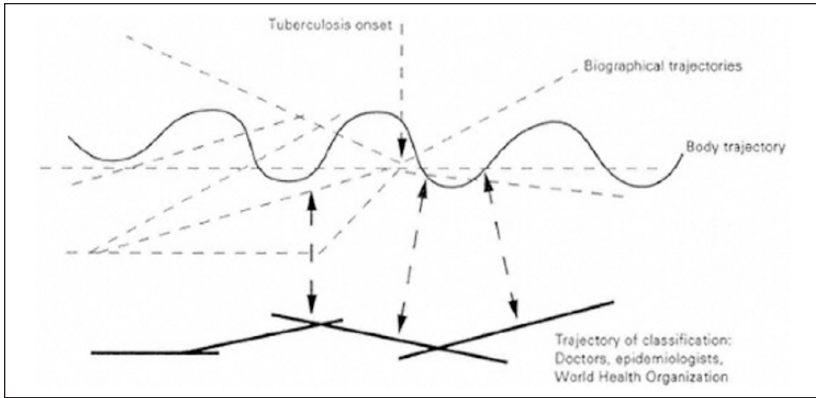


**Figure 1.** Strauss and Corbin’s body-biography chains in three different illness scenarios: chronic, catastrophic, and acute with recovery. Source. From Star and Bowker (1997). Note. The body and the self are connected but not always tightly coupled.

graphical representations shown in Figure 1 to represent the ways different kinds of illnesses can shift biographies, with chronic illness having a different shape than a sudden, catastrophic illness or an acute illness followed by a recovery.

Bowker and Star expand on this representation by adding a third trajectory: the trajectory of categories. Using the case of tuberculosis patients, they show the interplay of biography, body, and disease classification (Figure 2). Of particular interest is the influence of the latter. When a patient’s disease gets reclassified, regardless of the symptoms he is experiencing, this has an impact on biographical trajectory through increased or decreased restrictions on activity or new courses of treatment.

Similarly, NCLB classifications can become a source of torque in the student, teacher, and school trajectories over time. In one instance, New York City built the *Teacher Data Report (TDR)*, a rating of the city’s teachers based on test scores using NCLB categories. It did not take long for the *New*



**Figure 2.** The topology-typology twist shows tensions among bodies, biographies, and classification.

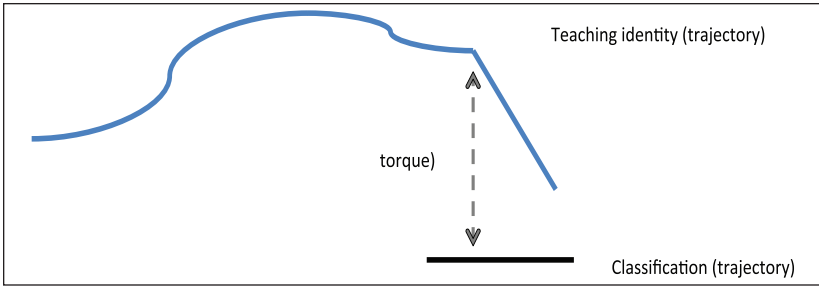
Source. From Star and Bowker (1997).

Note. The topology of the body-biography gets pulled by the trajectory of classifications, as people work to fit and resist their own experiences with that scheme. The dotted lines represent the negotiations.

*York Post* to stake out the house of Pascale Mauclair of P.S. 11, the teacher with the city’s lowest ranking, under the headline “Queens Parents Demand Answers Following Teacher’s Low Grade” alongside her picture and annual salary (Roberts, 2012). Outrage ensued. Bloggers quickly came to the teacher’s defense, quoting her colleagues and principal as deeming her “an excellent teacher” (Casey, 2012). The school’s principal, Anna Efkarpidis, was quoted as saying, “I would put my own children in her class.” As a post on the United Federation of Teachers *Edwize* blog reported,

Mauclair is an ESL [English as a second language] teacher, and over the last five years she has had small, self-contained classes of recently arrived immigrants who do not speak English. Her students arrive at different times of the school year, depending upon that date of their family’s migration; consequently, it is not unusual for her students to take the 6th grade exams when they have only been in her class for a matter of a few months. Two factors which produce particularly contorted TDR results—teaching the highest academic need students and having a small sample of students that take the standardized state exams—define her teaching situation. (Casey, 2012)

The classification of Mauclair’s teaching, which carried the weight of “truth” through “empirical” testing data, torqued her teaching identity. (See Figure 3 for a schematic representation.) Her colleagues and principal saw



**Figure 3.** A schematic representation of the torque in Pascale Maclair’s teaching identity trajectory upon the introduction of the classification scheme introduced by the TDR.

Note. The flat line at the bottom represents the introduction of the classification “worst teacher in New York City.” TDR = Teacher Data Report.

her as a “good” teacher, and yet this description of her collided with a headline on the front of a hometown newspaper identifying her as the city’s worst teacher, based on an NCLB classification scheme.

Stories of unmerited public shaming in the name of accountability abound, all with themes of the erasure of context and history through NCLB’s classificatory infrastructure. Aside from the potential validity threats to standardized measures coming from highly mobile students or students learning English, other distortions come from the previously described flaws in the measurement schemes themselves. I conceptualize these distortions as coming from *classificatory collisions*, which I identified throughout my data set. In this article, I examine the consequences of classificatory collisions not at the level of individual biographies but at the smaller grain size of teachers’ instructional decision making. The educational torque arising from the countervailing forces in NCLB’s *how*—narrow standardized testing working in conjunction with the indeterminate modes of instruction—derails the humanistic purposes of education while rendering problematic forms of triage as commonsense in teachers’ work.

## Background of Study

### Research Context

I came upon these classificatory collisions in the course of a larger research project looking at instructional improvement in mathematics in urban school districts. Since 2007, the Middle School Mathematics and the Institutional Setting of Teaching (MIST) project has investigated large-scale support of

teachers' development of equitable mathematics instruction. Our research team identified urban school districts investing in viable strategies to meet these goals. The project examines mechanisms for instructional improvement from the district central office to schools and classrooms, collecting a variety of qualitative and quantitative data to document this work. We chose District B as one of our partners because of its investments in high-quality mathematics curriculum and intensive teacher professional development.

District B serves approximately 80,000 diverse students, of whom approximately 50% identified as Latino, 25% African or African American, and 15% European American.<sup>2</sup> More than 25% of all students are classified as English language learners (ELLs). The majority of the students qualify for free or reduced-price lunch. District B's student achievement patterns in middle school mathematics look similar to those of many urban districts. On the annual state assessment for the year of this study, 50% of the African American students met mathematics standards, compared with 70% of the Latino students and 80% of the European American students. About 50% of the ELL students met standards in mathematics on these tests.

As a part of its efforts to improve mathematics achievement, District B invested in developing an interim assessment instrument and a curriculum framework, as well as providing teachers with collaborative time. The assessment instrument, which I call the ABC, was linked to the curriculum framework. Written by mathematics specialists in District B's Curriculum and Instruction department, this document specified which standards should be taught at each grade level, accompanied by an instructional pacing guide. The ABC provided accountability for the curricular framework by evaluating students on standards that corresponded with the previous six weeks' instruction. Mathematics coaches and teachers wrote the test items on the ABC over the course of several weekends. As a tool, the ABC assessments did not go through the same development process as the state test, lacking both field-testing and psychometric validation, yet they were highly consequential in practice. District personnel monitored schools through ABC reports.

*Creekside Middle School: Concerted compliance with NCLB.* Creekside Middle School caught my attention because of its concerted compliance with NCLB's accountability logics. A strong administrative charge to improve on AYP proficiency targets shaped the norms and values of Creekside's seventh-grade math teacher workgroup. Having worked as Creekside's principal for 5 years, Mr. Russell attended the workgroup's weekly meetings and often emphasized this goal. The accountability press extended beyond the workgroup meetings and into the broader school environment, from the introduction of all-day Saturday Math Camps to Mr. Russell's morning announcements reminding

students to have a “good breakfast” on testing days. Creekside employed a full-time data manager to produce and monitor reports of both the ABC and state test. Representations of data and students’ progress were displayed throughout the school, including on classroom walls labeled with children’s names. The reports broadcast their stories everywhere, underscoring a strong orientation toward their narratives of student learning.

The schools’ student and teacher demographics reflected those of the district, with a diverse group of students served by primarily European American teachers. Creekside’s students were roughly split in thirds among students identified as African American, European American, and Latino. About 60% of the students qualified for free or reduced-price lunch.

The teachers described Mr. Russell as highly involved in their classroom instruction. In interviews, they all reported that he observed them at least once during the school year, with the majority of these visits being informal. Formal observations were accompanied by written feedback, creating a sense of instructional accountability. Mr. Russell worked closely with the school’s mathematics coach Tiffany. Together, they reported identifying mathematics teachers in need of support, and these people became central to Coach Tiffany’s classroom-level work. As we found with other coaches in the MIST project, Coach Tiffany did not work with all mathematics teachers due to their own resistance, perceived need, or her limited time (Gibbons, 2012).

*Focal data and research approach.* Creekside’s seventh-grade math team consisted of eight mathematics teachers, a special education assistant, Coach Tiffany, and Principal Mr. Russell, who led the meeting. Both Mr. Russell and Coach Tiffany facilitated the weekly grade-level team meetings, which lasted about 35 min during the school day. The seventh-grade mathematics team had more experience ( $n = 8$ ,  $M = 5.3$ ,  $SD = 7.6$ ) than the department as a whole ( $n = 12$ ,  $M = 4.4$ ,  $SD = 6.8$ ). About two thirds of the teachers self-identified as White, whereas the other one third self-identified as African American, Latino, or Pacific Islander.

In interviews, teachers reported that these meetings usually focused on planning and reflecting on lessons, sharing resources, and looking at ABC data to talk about getting scores up. As our project sought to understand instructional sensemaking across the various settings of the district, we videotaped four of the team’s weekly team meetings to understand teacher workgroups’ contribution to professional learning over the course of the school year.

In my research, I look at data using precepts from ethnomethodology (Garfinkel, 1967) combined with methods from sociolinguistics (Saville-Troike, 2008) to understand meaning-making in conversation. To support this

work, I parse conversations into *episodes of pedagogical reasoning* (EPRs) to analyze how educators make sense of instructional problems (Horn, 2005). EPRs are moments in teachers' talk where they share an instructional decision and provide a rationale for making it. These episodes can be brief, such as a teacher saying, "I don't do that activity because that content is not on the test." The more interesting EPRs involve multiparty conversations, as these typically involve a negotiation of both practice and the reasons for it, revealing educators' sensemaking. In this analysis, I found moments of *classificatory collision* by identifying extended EPRs where educators' conversations treated problematic practices as commonsense and used accountability as a justification. By *problematic practices*, I was specifically interested in forms of instruction that stood to undermine the vision of humanistic and equitable mathematics education.

## Findings

In this section, I present three instances of teachers' instructional decision making characterized by classificatory collision and that can be viewed as educational triage. Although the educators leading the discussions had high commitments to their students' success, the classificatory infrastructure of NCLB permeated their work practice and led to educational torque. Through these examples, I show three things: (a) triage happens, (b) triage affects children who are objects of the activity as well as children who are not, and (c) the epistemology of testing has, in some instances, naturalized itself as a form of commonsense.

### *Excerpt 1: Naturalizing Categories, or Triage Happens*

In a meeting on January 26, 2012, Mr. Russell had the data manager list all the seventh-grade African American students alongside their scores from last year's state test and the three interim ABC assessments from the current academic year. Students' performances on each of the assessments were coded as either *growth*, *bubble*, *proficient*, or *commended*. These lists were then disaggregated by teacher, printed up, and distributed in the meeting. Mr. Russell asked the teachers to use their "perception data" alongside the "hard data" on the printouts to predict which of the "AA students" would pass the end-of-year test. Mr. Russell explained that this would help them identify bubble kids to target for "personalization" in the classroom, as well as extra tutoring and spring break math camp.

In the end, the teachers' "soft" perception data lent a more generous sorting of the students, generally moving students toward higher achievement

categories and putting more students in the pool of bubble students. Mr. Russell responded positively to the new sorting:

You guys think more kids you can get over the hump, and that's encouraging, very encouraging because I think the lowest number besides Mr. Floyd was somewhere in that 27% range and so on. So, you guys are saying, you know, that number's probably closer to 60, 65% of kids that we can go get either, you know, bubble, passing, or commended kids and that's encouraging news, especially with our most difficult sub-pop.

There are several points to be made here about NCLB as a classification system and its naturalization into practice. First, as Mr. Russell's words indicated, this activity was indeed an example of educational triage. Although large-scale investigations have alternately confirmed (Lauen & Gaddis, 2012) and disconfirmed (Springer, 2008) triage as a practice, this observational data show us that triage does in fact happen. Second, as a classification system, the epistemology of NCLB invites a stance of certainty. With the testing technologies of statistics, distributions, and cut points, the data reports carry a weight that invites a positivistic take on the stories they tell about teaching and learning (Horn, Kane & Wilson, 2015). Although the certainty of the test data was mitigated by the ambiguity of the teacher's "perception data," the role of the latter was mainly to provide slight shifts in the classification dictated by the former, not a radical reorganization of categories. The introduction of "perception data" could be taken as a way to humanize the test-based classification system; however, this attempt gets little traction as it collides with the hard-and-fast sanctions the school will face if they do not address the test-based sorting of their students.

Mr. Russell's last utterance about the "most difficult sub-pop" illustrates what is perhaps the most troubling naturalization of the classificatory infrastructure. NCLB's provision to hold schools accountable for the progress of subpopulations of students was, in many ways, its boldest innovation and its gesture toward ensuring the education of all students. In practice, it has had the uncomfortable effect of reifying existing narratives of race and achievement, what has been called "gap gazing" (Gutiérrez, 2008; Ladson-Billings, 2013; Milner, 2012), referring to the painting of static pictures of inequities in schools. Consonant with the way classification creates indeterminate pasts and creates standard narratives that appear universal, achievement gaps, as a narrative, become naturalized in schools through test-based accountability. Although Mr. Russell's intention in this activity is to focus teachers on what they can do for their "most difficult sub-pop," there is a paradoxical way in which the very act of disaggregation reinstates the unequal achievement of the school's African American students as inevitable.

The potential for equitable teaching was also limited by a concern about resources. Specifically, Mr. Russell said,

[W]e have to worry about resource allocation, let's make sure we're getting the right kid in Camp that's gonna behave correctly. I mean, I know that's difficult, but definitely something we have to focus on at a large school with limited resources.

This statement put a threshold on which bubble kids merited these scarce resources, charging the teachers to “make sure [they’re] getting the right kid in Camp that’s gonna behave correctly.” Because NCLB is a mostly unfunded mandate (Darling-Hammond, 2007), Mr. Russell’s concern about resources is understandable. At the same time, given the unequal scrutiny applied to the same student behavior for Black and White students (Finn & Servos, 2013), this threshold does not create richer instructional opportunities for all students—not even all African American “bubble kids”—bounding access to these resources based on teachers’ assessment of student behavior.

Notably, over the course of 27 min, the mostly White teachers did not take up what it would mean for them to motivate and individualize their instruction for their African American students, making this conversation colorblind (Yoon, 2012). The reinstatement of student deficit narratives and simultaneous glossing of teachers’ possible contributions to underachievement exemplifies the kind of everyday racism at work in schools and mathematics departments (Martin, 2009).

A humanistic response, in contrast, might dig deeper into the sources of the African American students’ underperformance, providing narratives of student competence, raising questions about the standardized testing measure, and supporting a critical reflection of teachers’ relationships with the children in their classes. Under the policy’s logics, however, this conversation provides no outward pushback, even among teachers who revealed in interviews a critical lens on the accountability regime.

### *Excerpt 2: Excellence for Some, or Leaving Children Behind*

In another meeting of the Creekside seventh-grade team on April 3, 2012, the state test was just weeks away. Coach Tiffany led the meeting to discuss a number of logistical issues, including an upcoming series of Saturday Schools and an all-day Math Camp, both forms of supplemental instruction for the bubble kids. Math Camp was to take place on a regular school day, but the attendees would be pulled out of all of their classes and only do math. This structural reorganization, clearly a form of educational triage, led to new dilemmas that Coach Tiffany brought up with the teachers:



So next week obviously, there'll be 2 days where we're kinda *without* our kids, some of us. Those of you that are staying behind, what do y'all want to teach the kids. Because it, it can be kind of like a, a mini-review for them, sort of like a mini-camp. So the kids that are gonna be left behind, you have to consider, are the really, really struggling children and then your kids that don't need camp. So you kinda have your, a big gap of kids that are staying.

Once again, the production of Math Camp and its related dilemmas are clearly an outcome of NCLB classifications. In the context of the policy with its emphasis on percentage proficiency, Math Camp and Saturday School for bubble students make perfect sense.

From the view of humanistic education, however, it is easy to see this as a misguided solution. As Gay (2007) has pointed out, NCLB has led to a situation wherein

important knowledge, attitudes, values and skills needed for learning how to be morally, ethically and emotionally healthy human beings, caring community members and worthy citizens in a culturally pluralistic society and world are being woefully neglected. (p. 291)

Math Camp, Saturday School, and other forms of triage work against these broad and varied notions of excellence within a humanistic education framework. Instead, these activities can be seen as a finer-grained system of tracking wherein the bubble students are visibly removed from their classrooms, leaving a drastically heterogeneous group (the "really struggling children" and the "kids that don't need camp") behind for a regular school day.

As with any form of tracking, status issues inevitably arise for students, especially when achievement levels are so publicly visible. The students who attended the Math Camp surely knew that they were being targeted for improvement. The students left behind could likewise figure out where they stood in the NCLB classificatory system. As Rosenholtz and Wilson (1980) demonstrated, students are usually able to size up ability grouping with high levels of agreement: They know who is "smart" and who is "dumb." Separating children in this obvious way reinforces academic status (E. G. Cohen, 1994; Horn, 2012) which in turn sets students' and teachers' expectations of achievement, often leading to self-fulfilling academic prophecies. Instead of positioning children as mathematically competent (Dunleavy, 2015; Gresalfi et al., 2009; Gutiérrez, 2013), this social arrangement further conflates test achievement with ability.

The radical heterogeneity of the students left behind also poses instructional challenges for the teacher. Because of the amplification of status problems, even a teacher highly skilled at differentiated instruction would

find an inordinate challenge in coming up with meaningful activities. Not only is the idea of individual excellence compromised, so too are meaningful instructional environments. The collision, in this case, comes from the exclusive focus on getting supplemental instruction for bubble kids and a disregard for the consequences to the remaining students and teachers.

### *Excerpt 3: Reducing Uncertainty, or Testing Trumps Teaching*

This final example comes from the same April 3, 2012 meeting described in Excerpt 2. Coach Tiffany and her colleagues pursued the idea of what to work on with the students left behind while the bubble students were at Math Camp. Then Tiffany thought that perhaps “more data” would help them identify useful topics to reteach on those days. She proposed asking Mr. Russell to let them “shut down” instruction to administer a full-day benchmark test that would yield data on what to work on. This plan would require having all their math classes devote an entire day to additional testing.

Floyd, a senior teacher, hesitated, and the following discussion ensued:

- 
- |   |          |   |
|---|----------|---|
| 1 | Floyd:   | I don't know the (3-s pause). Would it serve any purpose for us to gather any more data? Would it serve us to know any more than we already know? And are we willing to exchange instruction time for <i>shutting</i> down and, and testing again?  |
| 2 | Tiffany: | I think until we= <sup>3</sup>  |
| 3 | Erica:   | =I would rather, if we, if he would let us do it like in one day, I would rather that, that way, then we just lose one day.   |
| 4 | Tiffany: | Well yeah cause then it's a full test mode. I think until the kids sit through—because when they took the, what did the 8th graders, they Benchmarked in math when they took, the practice, the mock =  |
| 5 | Erica:   | =7th grade, I think yeah=   |
| 6 | Tiffany: | =that's what it was, the mock writing, okay?  |
| 7 | Regina:  | The Writing Benchmark.  |
| 8 | Tiffany: | Mhm. So until those kids can sit down, have the timer up on the board, only be given what they're gonna be given, learn how to ask for scratch paper, “Here's your timer, when it's done it's done, if you're bubbled in, great.” We had a lot of 8th graders that didn't bubble in, in time, and they learned, “I've really gotta bubble in as I go because I'm not gonna be done.” Until they've had that practice, they don't know how they're gonna do. (2-s pause). That's my opinion. |
-

Floyd questioned the premise of Tiffany's notion of exchanging instructional time for testing time. When he voiced his concerns in turn 1, his colleagues presented their rationale. First, Erica seemed to disagree that the loss of instructional time was substantial (turn 2, "just one day"). More expansively, Tiffany, with contributions from Regina and Erica, countered Floyd's concern by illustrating the value of "full test mode" (turn 4) using hypothetical student and teacher dialogue about the importance of correct bubbling to support her view that students did not know how to test well (turn 8).

Floyd questioned this argument, saying that Tiffany was describing test-taking skills. She pushed back once again and claimed that her proposal went beyond that: The data were important to show them what to reteach. Floyd continued his effort to challenge the proposed instructional shut down: "Are you gonna test *all* of the kids, even though you have kids within that group who you *know* what they know?"

Floyd claimed that he, as a teacher, knew what most of his students knew. Coach Tiffany, however, countered once again:

Yeah. The 8th grade teachers liked having that data because they said, "Okay on"—I think it was Pythagorean Theorem, 60% of the 8th graders did not come close to mastering that, where the teachers thought they had, because in class they were doing it (*snaps finger*). They thought they knew it, they didn't know word problem or on that test they bombed it and so they were like, "Okay, we really, we thought we were doing really good on this, we need to back it up." And we added that to our Camp, so it was, for them it helped them see, "Okay we thought they had this mastered, they have no clue what this is." They *didn't* think they had mastered volume and they all did really good on volume. So it was, it was good for them, kind of an eye-opener.

In relating the story of the eighth-grade teachers' data-induced "eye-opener," Tiffany communicated that the tests' knowledge of student learning was more accurate than the teachers' knowledge of student learning. The eighth-grade teachers' misjudgments went in both directions: The students did *not* know the Pythagorean theorem, and they really *did* know volume. Tiffany's story became an argument for the epistemology of testing over the epistemology of teaching, with the certainty of the data beating out the ambiguity of the classroom.

Tiffany and her other colleagues' enthusiasm for shutting down instruction for testing over signals a triumph in this setting of NCLB's relentless pursuit of hard evidence. To be sure, effective instruction requires knowledge of what learners know. Yet in the story about the eighth-grade teachers' inaccurate inference of their students' geometry knowledge, the test measures

themselves remained opaque. The misalignment in judgment is assumed to be entirely because of their misreading and not because of a flaw in the test—questions that were either too hard or too easy, miscues, or confusing distracters. The testing technology was not doubted, and neither were the narratives of students knowing and not knowing it produced. In this case, teachers' professional judgment about students' understanding collided with a deep seated faith in standardized testing instruments.

## Discussion and Conclusion

NCLB provides a classificatory infrastructure rooted in techno-rational discourses to teaching and moves instructional decision-making away from the values of humanistic education—a phenomenon I call torque. Rather than a direct line from the threat of sanctions to deliberate instructional improvement, the requirement for immediate and measurable growth on narrow (and often problematic) instruments, coupled with undefined instructional values, encourages the triage documented here to meet the policy's goals. In fact, when proficiency rates are the target, triage is a sensible response, as it optimizes resources and effort while focusing squarely on the desired outcome.

The classificatory infrastructure of NCLB leads to torque in the lives of students, teachers, and schools—an especially burdensome consequence for urban schools. As the Creekside educators worked deliberately and arduously within the logics of the policy, uncomfortable things happened. The principal narrated African American students as the “most difficult sub-pop,” reinforcing negative academic stereotypes, in his efforts to support their education, whereas avoiding the harder, more uncertain discussion about potential reasons for students' underachievement. Even the supplemental education offered was limited to “well-behaved” students, a fraught category in most schools when it comes to Black and Latino students, whose behavior is often unfairly scrutinized compared to that of White and Asian students (Finn & Servos, 2013). An instructional coach orchestrated a scheme to actually leave some children behind while others—the “bubble kids”—got intensive mathematics instruction, a situation that made academic status highly visible in the school, undermining richer and more varied notions of excellence. The coach and most of her colleagues decided to favor testing time over instructional time, invoking the certainty of test data over the ambiguity of classroom assessments

From a *policy as learning design* perspective, the torque came out of tensions in the *how* of meeting the policy's requirements, rendering these problematic teaching practices commonsensical. The demand to increase test scores foreclosed any attempts to bring in other visions of education into the

teachers' decision making, as happened in Excerpt 3 when Floyd suggested additional instruction would contribute to student learning more than additional testing. His appeal for humanistic values could not overcome the techno-rational appeal of testing.

The educational research literature offers examples of places that resist torque and adhere to humanistic values, despite policy pressures to do otherwise. Maintaining a commitment to equity in the humanistic sense lead teachers to work around bureaucratic structures, as they prioritize students and their learning over rules and regulations—what Sandra Stein (2002) called “thoughtful noncompliance” in her study of Title 1 implementation in urban schools. Responsiveness and thoughtful noncompliance place high demands on teachers' professional judgment as they make sense of their instructional decisions day-to-day.

In the case of Creekside, torque came from dutifully meeting the policy demands, as teachers negotiated instructional decisions. In these excerpts, I showed existing stereotypes being reinstated (Excerpt 1), academic achievement getting narrowly reified (Excerpt 2), and valuable instructional time being traded in for additional testing (Excerpt 3). NCLB, as a policy, mediated these educators' activity and rendered it sensible.

Although these instances of torque may seem surprising to some, the timeline and underlying assumptions of NCLB play into these events. First, the demand for immediate evidence of improvement may work against teachers' development. Responsive and rich instruction, a possible endpoint for improvement, is a particularly complex and rare form of teaching, demanding the marshaling of expertise, learning over time, and challenging interpretive work. In aiming for this kind of instruction, improvement may involve an initial backslide, as teachers grapple with new practices—a backslide that may not be deemed worth the risk, as it would be brought to light through the surveillance mechanisms of NCLB. Indeed, more often, schools under threat of AYP sanctions are more likely to retreat to narrow curriculum (Au, 2007; Darling-Hammond, 2007) and less meaning-centered forms of mathematics instruction (Ellis, 2008), to the possible detriment of students' sense of mathematical competence (Gutiérrez, 2013; Horn, 2012). In addition, the epistemology of NCLB, with its positivistic overtones of certainty, works against the epistemology of responsive teaching, which is more ambiguous and situated with the complex ecology of the classroom. In light of these obstacles, the teachers' workarounds make sense.

Returning to Bowker and Star (1999), the authors caution against classification systems that favor epistemologies of certainty. In discussing the example of apartheid, for instance, they point out that a “purified system of categories” yields “many ironies and much individual suffering” (p. 219).

Sorting the complex topography of student learning into the four categories of *growth*, *bubble*, *proficient*, and *commended* certainly carries with it the hope of such purity—and, along with it, ironies and suffering. There is no commendation, for instance, for the student who enters Creekside achieving two grade levels below his age and who, by his own hard work and dedication of his teachers, manages to end the year only half a grade behind. He still would not be viewed as *proficient* despite this enormous accomplishment: the category of *growth* collides with the reality of his tremendous achievement, torquing the narrative of what was taught and learned. The instruments are a technology, and an imperfect one at that. Yet the techno-rational appeal of data turns the interpretive work of teaching into a naively positivistic one, which, in many ways, is the hope of the policy itself.

A simplistic reading of this article would lead to the conclusion that classification is inherently bad or accountability is untenable. Neither is a foregone conclusion here. Classification is inevitable, but systems need to provide humane treatment of individuals through nuance and feedback into the mechanisms of classification themselves. Pascale Maclair, in other words, should not have been deemed “the worst teacher” when plenty of evidence indicated otherwise. The hypothetical student whose prodigious learning goes undetected by standardized testing should have the means to be recognized. However, at Creekside, as in other sites under the thumb of NCLB, the instruments’ decree is final and uncontestable. Alternative ways of showing competence or learning are not built into the classificatory infrastructure, let alone modes of critique and feedback for the means of classification, other than the continual shifting (or, one might say, gaming) of proficiency cut points. The hope of this analysis is to highlight the unintended consequences of the current form of school accountability as they play out in educators’ pedagogical reasoning. In this way, we can learn from it and design systems less fraught with contradictions to better support the development of urban students, teachers, and schools.

### **Acknowledgment**

I thank my research team, Mollie Appelgate, Jason Brasel, Brette Garner, Britnie Kane, and Jonee Wilson for their assistance in data collection and analysis, along with the whole MIST team. Special thanks to the educators who allowed us in their schools at a time of heightened professional scrutiny. Finally, I am grateful for the Spencer Foundation’s convenings of the *Evidence for the Classroom* grant recipients and the tremendous intellectual community they provided.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by grants from the National Science Foundation under Grant No. DRL- 1119122 and the Spencer Foundation.

## Notes

1. Although the focal school was beholden to No Child Left Behind (NCLB), the local education authority also had linked standardized test scores to teacher evaluations, similar to the provisions in Race to the Top (2009), the accountability policy that followed NCLB. Although I limit my discussion to NCLB, many of these contradictions are likely to be found in similarly formulated accountability policies.
2. All names are pseudonyms. Numbers are rounded to prevent reverse look up of research sites.
3. The = in the transcripts indicates *latching*, or the quick succession of turns that signals speakers building off each other's ideas.

## References

- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258-267.
- Au, W. (2013). Hiding behind high-stakes testing: Meritocracy, objectivity and inequality in U.S. education. *International Education Journal: Comparative Perspectives*, 12(2), 7-19.
- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary mathematics. *Elementary School Journal*, 93, 373-397.
- Battey, D. (2013). "Good" mathematics teaching for students of color and those in poverty: The importance of relational interactions within instruction. *Educational Studies in Mathematics*, 82, 125-144.
- Beachum, F. D., Dentith, A. M., McCray, C. R., & Boyle, T. M. (2008). Havens of hope or the killing fields the paradox of leadership, pedagogy, and relationships in an Urban Middle School. *Urban Education*, 43, 189-215.
- Boaler, J., & Greeno, J. G. (2000). Identity, agency, and knowing in mathematics worlds. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 171-200). Westport, CT: Ablex Publishing.
- Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *The Teachers College Record*, 110, 608-645.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Bruner, J. S. (1961). *After John Dewey, what?* New York, NY: Bank Street College of Education.
- Casey, L. (2012, March 1). *The true story of Pascale Maclair*. Retrieved from <http://newpol.org/content/true-story-pascale-maclair>

- Cobb, P., & Jackson, K. (2012). Analyzing educational policies: A learning design perspective. *Journal of the Learning Sciences, 21*(4), 487-521.
- Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 25-46). New York: State University of New York Press.
- Cohen, D. K. (2011). *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.
- Cohen, E. G. (1994). *Designing groupwork: Strategies for the heterogeneous classroom* (2nd ed.). New York, NY: Teachers College Press.
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of "No Child Left Behind." *Race Ethnicity and Education, 10*, 245-260.
- Diamond, J. B., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. *Yearbook of the National Society for the Study of Education, 106*, 241-263.
- Dunleavy, T. K. (2015). Delegating Mathematical Authority as a Means to Strive Toward Equity. *Journal of Urban Mathematics Education, 8*(1), 62-82.
- Ellis, M. (2008). Leaving no child behind yet allowing none too far ahead: Ensuring (in)equity in mathematics education through the science of measurement and instruction. *The Teachers College Record, 110*, 1330-1356.
- Finn, J. D., & Servos, T. J. (2013). Misbehavior, suspensions, and security measures in high school: Racial/ethnic and gender differences. In D. Losen (Ed.), *Closing the school discipline gap: Research for policymakers* (pp. 44-58). New York, NY: Teachers College Press.
- Gándara, P., & Baca, G. (2008). NCLB and California's English language learners: The perfect storm. *Language Policy, 7*, 201-216.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
- Gay, G. (2007). The rhetoric and reality of NCLB. *Race Ethnicity and Education, 10*, 279-293.
- Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. New York, NY: Teachers College Press.
- Gibbons, L. K. (2012). *Examining mathematics coaches in supporting teachers to develop ambitious instructional practices: A three-paper dissertation investigating context, knowledge, and practice* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.
- Gresalfi, M., Martin, T., Hand, V., & Greeno, J. (2009). Constructing competence: An analysis of student participation in the activity systems of mathematics classrooms. *Educational Studies in Mathematics, 70*, 49-70.
- Gutiérrez, R. (1996). Practices, beliefs and cultures of high school mathematics departments: Understanding their influence on student advancement. *Journal of Curriculum Studies, 28*, 495-529.
- Gutiérrez, R. (2008). A "gap gazing" fetish in mathematics education? Problematising research on the achievement gap. *Journal of Research on Mathematics Education, 39*, 357-364.



- Gutiérrez, R. (2013). The sociopolitical turn in mathematics education. *Journal for Research in Mathematics Education*, 44, 37-68.
- Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education*, 34, 37-73.
- Ho, A. D. (2008). The problem with “proficiency”: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, 37, 351-360.
- Horn, I. S. (2005). Learning on the job: A situated account of teacher learning in high school mathematics departments. *Cognition & Instruction*, 23(2), 207-236.
- Horn, I. S. (2007). Fast kids, slow kids, lazy kids: Framing the mismatch problem in math teachers’ conversations. *Journal of the Learning Sciences*, 16(1), 37-79.
- Horn, I. S. (2012). *Strength in numbers: Collaborative learning in secondary mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Horn, I. S., Kane, B. D., & Wilson, J. (2015). *Making Sense of Student Performance Data: Mathematics Teachers’ Learning through Conversations with Assessments*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Ingersoll, R. M., & Perda, D. (2010). Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal*, 47, 563-594.
- Jencks, C., & Phillips, M. (Eds.). (2011). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32, 465-491.
- Ladson-Billings, G. (2011). Yes, but how do we do it? Practicing culturally relevant pedagogy. In J. Landsman & C. Lewis (Eds.), *White teachers/diverse classrooms: Creating inclusive schools, building on students’ diversity, and providing true educational equity* (2nd ed., pp. 33-46). Sterling, VA: Stylus.
- Ladson-Billings, G. (2013). “Stakes is high”: Educating new century students. *The Journal of Negro Education*, 82, 105-110.
- Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB’s subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis*, 34, 185-208.
- Leonardo, Z. (2007). The war on schools: NCLB, nation creation and the educational construction of whiteness. *Race Ethnicity and Education*, 10, 261-278.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.
- Lortie, D. C. (1975). *Schoolteacher: A sociological inquiry*. Chicago, IL: University of Chicago Press.
- Loveless, T. (2013). *The 2013 Brown Center report on American education: How well are American students learning?* (Vol. 3, No. 2). Washington, DC: Brookings. Retrieved from <http://www.brookings.edu/2013-brown-center-report>
- Malen, B. (2011). An enduring issue: The relationship between political democracy and educational effectiveness. In D. Mitchell, R. Crowson, & D. Shipps (Eds.), *Shaping education policy: Power and process* (pp. 23-60). New York, NY: Routledge.
- Martin, D. B. (2009). Researching race in mathematics education. *The Teachers College Record*, 111, 295-338.

- McGee, E. O., & Martin, D. B. (2011). "You would not believe what I have to go through to prove my intellectual value!" Stereotype management among academically successful Black mathematics and engineering students. *American Educational Research Journal*, *48*, 1347-1389.
- Milner, H. R. (2010). *Start where you are, but don't stay there: Understanding diversity, opportunity gaps, and teaching in today's classrooms*. Cambridge, MA: Harvard Education Press.
- Milner, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, *43*, 693-718. doi:10.1177/0021934712442539
- Moses, R. P., & Cobb, C. E. (2001). *Radical equations: Civil rights from Mississippi to the Algebra project*. Boston, MA: Beacon Press.
- Nasir, N. I. S., & Shah, N. (2011). On defense: African American males making sense of racialized narratives in mathematics education. *Journal of African American Males in Education*, *2*, 24-45.
- Roberts, G. (2012, February 26). "Queens parents demand answers following teacher's low grade." Retrieved from [http://www.nypost.com/p/news/local/cursed\\_with\\_the\\_worst\\_in\\_queens\\_f5wLhEdDRN1W19h1GQgxAM](http://www.nypost.com/p/news/local/cursed_with_the_worst_in_queens_f5wLhEdDRN1W19h1GQgxAM)
- Rosenholtz, S. J., & Wilson, B. (1980). The effect of classroom structure on shared perceptions of ability. *American Educational Research Journal*, *17*(1), 75-82.
- Saville-Troike, M. (2003). *The ethnography of communication: An introduction* (3rd ed.). Malden, MA: Blackwell Publishing.
- Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, *18*, 253-286.
- Spillane, J. P., Reiser, B. J., & Gomez, L. M. (2006). Policy implementation and cognition In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 47-64). New York: State University of New York Press
- Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*, 556-563.
- Star, S. L., & Bowker, G. C. (1997). Of lung and lungers. In A. C. Strauss & J. M. Corbin (Eds.), *Grounded theory in practice*. (pp. 197-228). Thousand Oaks, CA: Sage.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629.
- Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis*, *25*, 423-448.
- Stein, S. J. (2002). *The culture of education policy*. New York, NY: Teachers College Press.
- Stoskopf, A. (1999). The forgotten history of eugenics. *Rethinking Schools*, *13*(3), 12-13.
- Strauss, A. C., & Corbin, J. M. (1997). *Grounded theory in practice*. Thousand Oaks, CA: Sage.
- Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.
- Tate, W. F. (1995). Returning to the root: A culturally relevant approach to mathematics pedagogy. *Theory Into Practice*, *34*, 166-173.

- Trujillo, T. M. (2013). The disproportionate erosion of local control urban school boards, high-stakes accountability, and democracy. *Educational Policy, 27*, 334-359.
- U.S. Congress. (1964). *Elementary and Secondary Education Act*. Washington, DC: Author.
- U.S. Congress. (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.
- U.S. Department of Education. (2009). *Race to the top executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Ware, L. (2002). Race and urban space: Hypersegregated housing patterns and the failure of school desegregation. In *Widener Law Symposium Journal, 9*, 55-71.
- Wilson, J., Nazemi, M., & Jackson, K. (in preparation). *Learning from teaching: Identifying mathematics teaching practices in classrooms that support African American students*.
- Yoon, I. H. (2012). The paradoxical nature of whiteness-at-work in the daily life of schools and teacher communities. *Race Ethnicity and Education, 15*, 587-613.

### **Author Biography**

**Ilana Seidel Horn** is professor of Mathematics Education in the Department of Teaching and Learning, Vanderbilt University Peabody College. Her interests are in mathematics teachers' workplace learning and its relationship to equitable teaching practices.