

1

Introduction

The Accountability Debate in Texas: Continuing the Conversation

ANGELA VALENZUELA

The alleged “Texas Miracle” in education (Haney, 2000, 2001), combined with the 2002 reauthorization of the federal Elementary and Secondary Education Act (popularly known as the “No Child Left Behind Act” [NCLB]), has shifted what was once an intrastate debate over educational accountability to a national-level issue. Supporters of accountability claim that it promotes equity by making schools teach poor and minority children who have been historically neglected by our public school system (see Scheurich & Skrla, 2001; Scheurich, Skrla, & Johnson, 2000; Skrla, Scheurich, & Johnson, 2000a, 2000b). Opponents, including the contributors to this volume, argue that the Texas system of educational accountability has failed—and will continue to fail—Latina/o and other minority youth and their communities.¹ We interpret Texas-style accountability as exacerbating historic inequities, mainly through the collateral effects of state policy, but also through a systemwide failure to accommodate the needs and abilities of English-language learners (see the chapters by Alamillo, Palmer, Viramontes, & García; and by Ruiz de Velasco). Moreover, as McNeil points out (see chapter 3), the dramatic educational improvement attributed to Texas’ system of accountability is itself questionable. The state’s methods of collecting and reporting educational data, including the critically important high-stakes test scores, hide as much as they reveal. When the focus is shifted to Texas’ students’ performance on nationwide tests such as the American College Test (ACT) and Scholastic Assessment Test (SAT) 1, or when skyrocketing dropout and projected retention rates are factored in (see McNeil and Valencia & Villarreal), the state’s “miracle” looks more like a mirage.

That schools should be held accountable is indisputable. This volume does not suggest otherwise. Rather, what we question is the Texas model of accountability. Specifically, the authors reject high-stakes testing, the system's centerpiece. We further contend that the Texas approach is deeply flawed, for three interrelated reasons: for attaching high-stakes consequences—in the areas of retention, promotion, and graduation—to a single measure of students' academic abilities; for attaching high-stakes consequences to schools and districts and thereby encouraging a reductionist, test-driven curriculum; and for promoting a uniform and objectivist way of knowing, to the detriment of other cultures, languages, and approaches to knowledge.

Our collective admonition to the nation is that policies supporting high-stakes testing are harmful to all children, especially for children from poor, minority, or non-English-speaking families.² Indeed, these policies curtail or compromise the very achievement the public seeks. Moreover, state policies that attach high-stakes consequences to children's test scores are inherently invalid, undemocratic, and unjust (Heubert & Hauser, 1999). They distort the process of schooling, as well, through the creation of perverse incentives to "lose" children or limit curriculum, or both (see the chapters by Alamillo, Palmer, Viramontes, & García; Hampton; McNeil; Sloan; and Valencia & Villarreal). Finally, when the test is the sole or primary arbiter in decisions with such long-lasting consequences for children, we insist that students have a right to be assessed in a complete and fair manner, using as many criteria as may reasonably indicate children's cognitive abilities and potential.³

We would like to see the terms of the current debate over educational accountability overhauled. At issue is not whether schools and districts should be accountable, but what means should be used to accomplish the widely shared goal of ensuring that all children receive a high-quality education. When we allow the state to equate academic excellence with a single test score, when we agree to tie our children's performance on one test to their classroom teachers' jobs and school administrators' bonuses, we implicitly validate the host of values, presuppositions, and attitudes that underlie a flawed version of accountability. Recasting the debate, we hope, will draw necessary attention to the questionable nature of such typically unexamined assumptions.

In calling for a new approach to the ongoing conversation about accountability, we seek to create a larger public space for a Latina/o, research-based perspective and epistemology (see Padilla) in the development of a more just assessment system. In this space, we also see assessment as a tool that can remove schools from public scrutiny with the use of alienating,

technical language. In terms of how an “appropriate” or “necessary” educational experience gets defined within a system in which testing drives curriculum, prescribed notions of assessment may also contribute to subtractive cultural assimilation and thereby harm, control, and exclude children and their communities (see Alamillo et al., Hampton, Padilla, and Sloan). At its best, however, assessment affirms a core democratic principle that education is about shared governance and growing healthy children, communities, and citizens for a democracy.

Accordingly, what follows is a description of the current public policy landscape in Texas, providing the backdrop against which my own and others’ efforts to pass multiple assessment criteria legislation have taken place. The purpose of this narrative is to situate our modest but significant policy efforts to reclaim education as a publicly shared institution essential to active, democratic participation. This volume does not address such issues as test theory, measurement, or frameworks for assessment in any depth. Abundant scholarship already exists in these areas (e.g., Darling-Hammond, Aness, & Falk, 1995; Newmann, 1996; Meier, 1995, 2002; Valencia & Suzuki, 2001). Except for the contribution by Alamillo et al., this book engages Texas’ unfolding policy debate on accountability and its implementation. This introductory chapter provides an overview of the contributions to the volume. Much as my assessment of the struggle in Texas for multiple-criteria legislation is shaped by my position as a parent of two elementary-school-age children, Latina academic researcher, and community activist, so too do the other authors contribute with their unique experiences and perspectives.

THE UNFOLDING DEBATE ON ACCOUNTABILITY

For reasons that go beyond the scope of this chapter, discussions about educational policy issues, whether they take place among researchers or between researchers and the lay public, often are marred by misunderstandings. Each side unwittingly generates opposition to its views by failing to meaningfully engage challengers’ assumptions and perspectives. Specialized language and the tendency to assign multiple meanings to the same term (such as “standards-based reform”) only increase the levels of complexity and confusion. As researchers, we are typically somewhat removed from public life; we tend not to participate in the kind of debates and discussions that not only would teach us to think through our evidence and assumptions, but also force us to improve how we articulate our views and positions. My experience as one of a community of schol-

ars engaged in multi-voiced conversations about accountability has prompted me to try to articulate a new, more constructive frame for thinking about and taking action regarding the conceptual and rhetorical quagmire commonly referred to as “accountability.”

Background

My expertise on the state’s educational system is derived from research I have undertaken in Houston’s inner-city public schools over the past 12 years, first as a research associate at the Rice University Center for Education in Houston and now as a faculty member in the Department of Curriculum and Instruction and in the Center for Mexican American Studies at the University of Texas at Austin. While at Rice, I conducted, from 1992 to 1995, an in-depth case study of a large, overcrowded, predominantly Mexican, segregated high school. Most of the findings from this study are presented in my book, *Subtractive Schooling: U.S.-Mexican Youth and the Politics of Caring* (1999).

Subtractive Schooling critiques schools and state education policy as culturally and linguistically subtractive. Rather than building on the assets that children bring with them to schools, public education in Texas subtracts students’ linguistic, cultural, and community-based identities, to their academic and personal detriment. My more recent work links Texas’ system of testing to this practice of subtractive cultural assimilation (McNeil & Valenzuela, 2001; Valencia, Valenzuela, Sloan, & Foley, 2001; Valenzuela, 2000, 2002). A core idea here is that the state’s testing system is embedded within the larger edifice of “accountability,” and that it is this overarching framework that perpetuates a subtractive approach to the education of racial, cultural, and linguistic minorities. Children, along with their parents and communities, are treated as objects. The very notion of a mainstream, standardized educational experience implies a systemic disregard of children’s personal, cultural, and community-based identities. Rather than providing children with an empowering sense of how their lives can connect productively to the world that they inhabit, a test-centric curriculum compelled by the long arm of the state through standardized, high-stakes testing reduces children’s worth to their test scores.

Taking Action

In 2000, I joined two other contributors to this volume (Linda McNeil and Richard Valencia) to testify on behalf of plaintiffs (represented by the

Mexican American Legal Defense and Education Fund [MALDEF] in a federal court trial, *GI Forum et al. v. Texas Education Agency et al.* (2000). The focus of the suit was the adverse impact of the testing system in Texas: 87 percent of all students who fail the state's high school exit exam are either African American or Latina/o.⁴ All nine Latino and African American plaintiffs in the case were high school students who had been denied diplomas because they had failed to pass at least some portion of the reading, writing, or mathematics segments of the Texas Assessment of Academic Skills (TAAS) tenth-grade exit exam.⁵ Based on both their grades and completion of required course credits for graduation, all of the plaintiffs would have graduated, were it not for their performance on the TAAS test (Kauffman, 2000; Saucedo, 2000). In an ironic turn, MALDEF won the adverse impact argument but lost the case. Judge Edward Prado ruled that the harm against the students did not reach a "constitutional level" because the state had followed due process in the design and implementation of the test.⁶

The trial, and MALDEF lead attorney Al Kauffman's efforts, in particular, helped raise the issue of high-stakes testing to a higher level of public awareness in Texas. Following the trial, a small, multiethnic group of university faculty, graduate students, and grassroots advocates for children (Parents United to Reform TAAS Testing) united into a coalition working toward twin goals: to increase awareness of the harmful effects of high-stakes testing and to promote the use of multiple compensatory criteria for academic assessment (see the chapter by Valencia & Villarreal; and Valenzuela 2002). Through our published work, as well as through opinion-editorial pieces, scholarly conferences, community presentations, and the dissemination of pertinent information through listservs and websites (especially www.texas-testing.org), we have been able to reach a variety of audiences. One especially successful effort to educate communities of color and other stakeholders throughout the state and nation took place during the opening days of the Texas legislative session on January 26, 2001. The event brought nearly all the contributors to this volume together with legislators, legislative staff, and an 800-strong, statewide representation of Latino students, parents, educators, and community activists. During the open-mike session, researchers, community members, and Latino educational leaders repeatedly voiced concern over test abuse. This public exchange, along with the federal trial and a rally organized by League of United Latin American Citizens (LULAC) activists in Houston on January 27, 2001 to challenge the accountability system, were among the first broad-based expressions of disenchantment in the Latino community regarding high-stakes testing.

From the start, our coalition also undertook action at the legislative level, joining forces with MALDEF, LULAC, and the Intercultural Development Research Association, a nonprofit, education advocacy organization for Texas' Mexican American community. When State Representative Dora Olivo (D-Rosenberg) indicated her willingness to sponsor our legislation during the 77th biennial meeting (2001) of the Texas legislature, we began working for the passage of HB 2118 and HB 2570. These bills called for the use of multiple compensatory criteria at the tenth-grade exit level, as well as at the third-, fifth-, and eighth-grade levels affected by SB 4, the state's new policy on social promotion (see the chapter by Valencia & Villarreal for details on SB 4's provisions and effects). SB 4, passed in 1999 when George W. Bush was governor, phases in new promotion gates, beginning in 2002–2003, with the state's third-grade cohort. Under this law, promotion to the fourth grade hinges heavily on the student passing the reading section of the state-mandated third-grade test, regardless of their overall academic accomplishments. The law also stipulates that promotion gates at the fifth- and eighth-grade levels be in place by 2007–2008.⁷

HB 2118 and HB 2570 explicitly addressed the need for *multiple alternative* assessment tools by proposing provisions to allow teacher recommendation, average grades in core subjects, and other test score information to compensate for poor performance on the Texas Assessment of Knowledge and Skills (TAKS) test, a newly devised and approved replacement for the TAAS. The state claims that a multiple-criteria system is already in place: Students must pass the TAAS/TAKS; they must maintain a 70 grade-point average; they must accumulate a specific number of credits for graduation; and they must attend school a certain number of days annually. However, since the decisive hurdle for graduation or promotion to the next grade level remains the student's TAAS/TAKS test score, what the state actually has in place is a multiple conjunctive criteria system, not a compensatory system. A student who met all criteria for graduation but failed the TAAS/TAKS would not graduate, because none of the other required criteria are allowed to compensate for poor test performance (Heubert & Hauser, 1999; Valencia & Bernal, 2000; Valenzuela 2002). As the plaintiffs in *GI Forum et al. v. Texas Education Agency et al.* (2000) remind us, far from being an extreme hypothetical scenario, this situation is painfully real for many Texas high school seniors.

The analogue we frequently offer is the college admissions process. Test scores typically are not the sole arbiter of college admittance. Most colleges assess applicants using multiple sources of information, often with the explicit intent of avoiding the possibility of losing otherwise excellent candidates because of their low scores on a single measure such

as the SAT. Moreover, decades of research on college testing shows that multiple criteria and a sliding scale of test scores and grades results in more valid decisions (Linn, 1982; Willingham, Lewis, Morgan, & Ramist, 1990). Decisions based on multiple criteria rather than on test scores in isolation also have been shown to have a smaller adverse impact on minorities and women (Haney, Madaus, & Lyons, 1993).

HB 2570 cleared its first major hurdle when it passed in the House with a majority vote; it did not receive a hearing in the Senate, however. HB 2118, on the other hand, never saw the light of day in either chamber. On the day that Representative Olivo was scheduled to testify on HB 2570, that fellow legislators told her as she was making her way toward the House floor that they had received word from the office of President George W. Bush not to support the legislation. As a result, many were reluctant to vote in favor of the bill. With hindsight, it is easy to see why the president's office would have taken such an active interest in HB 2570 and HB 2118. The federal NCLB legislation was in the process of being developed in Congress. A perceived substantive shift in "accountability" in Texas would have potentially made the president look weak at the national level.

We never viewed the legislation Representative Olivo sponsored as either partisan or as weakening President Bush's national goals for improving education. To the contrary, we believed—and still believe—that a fair assessment system could form the backbone of a more just and valid accountability system (Valenzuela, 2002). What this first foray into legislative action taught us (apart from the lesson regarding the power of political ties) was that we had failed to appreciate how many legislators fully equated "testing" with "accountability." To them, our anti-high-stakes-testing stance constituted an attack on the accountability system itself (see McNeil, 2000, for an excellent historical discussion of the evolution of accountability in Texas). For instance, when we testified on HB 2570 before the House Committee on Public Education on March 27, 2001, a lobbyist from the Texas Federation of Teachers expressed opposition to the legislation (www.house.state.tx.us/committees/audio77/400.htm). Claiming his organization's support for the development of the accountability system, he argued that the legislation weakened it. He felt that the legislation would nullify the driving force for change that a single-indicator system represents. These lessons have left us wiser and undeterred in our overall effort. HB 2570 and HB 2118 were refiled for the 78th legislative session in 2003, bearing new numbers (HB 336 and HB 337, respectively).

Implementation of the state's new TAKS test, administered on March 4, 2003, gave our efforts a special urgency. The TAKS is an espe-

cially troubling instrument, not only because it is a longer and more difficult exam than the TAAS, but also because its implementation collided in time with the enforcement of the state's new anti-social promotion policy (the centerpiece of SB 4; see Valencia & Villarreal, this volume). Using data provided by the State Board of Education (Texas State Board of Education, 2002), the state had estimated that 42,000 third-grade first-time test takers would fail the TAKS. It was our hope that the prospect of such high failure rates (and the possible retention in grade of so many children) would combine with our collective efforts to educate the public sufficiently to create an opening for the passage of Representative Olivo's bills. Before addressing the fate of the bills in the 2003 session, I first need to mention the results of the first TAKS administration.

The passage rate on the third-grade English and Spanish versions of the reading exam in the first administration were 89 percent and 82 percent, respectively (Texas Education Agency, 2003a). Numerically, this meant that "only" 32,659 students failed the exam. In state newspapers, the passing rate was hailed in positive, if not glowing, terms (e.g., Blackwell, 2003). A month later, on April 18, 2003, the editors of the *Houston Chronicle* argued explicitly against the Olivo legislation, citing the need for an accountability system that relies on "objective" test score data for determining "academic mastery." They further suggested that despite the "dire predictions" for third graders, the test results were "fairly good." To my knowledge, they were never charged with displaying a "soft bigotry of low expectations," one of President Bush's favorite campaign slogans against the critics of the Texas accountability system.

Privately, I wondered how a thirty-two-thousand-plus failure rate could be spun into a victory for the state. Were their expectations even more dire than the actual outcome? Should not the fact of tens of thousands of young children adversely affected be sufficient for us to take pause, particularly in a state that takes pride in the great strides it has allegedly made? As Valencia and Villarreal's (this volume) analyses demonstrate, failure rates are not only high—especially when considering their impact on retention and an enhanced probability of dropping out—they also obscure important subgroup variation that reveals a disproportionate impact of the system on Latina/o and African American youth. I also wonder why the burden was on us to demonstrate that *more* information should be used when judging children on decisions of such long-lasting consequence to them. The burden should be on the state to demonstrate that *less* information is better and that current failure rates are adequate, reasonable, and acceptable.

Representative Olivo was able to get nearly enough Republican and Democrat sponsors in the House to pass HB 336 and HB 337. However, the legislation never came out of committee (see the final chapter in this volume for a comprehensive analysis of the conservative legislative climate in the 2003 session). Olivo quickly adopted a new strategy to salvage the content of HB 336 and 337. She successfully amended a dropout prevention bill that the Chairman on the Committee on Public Education, Kent Grusendorf, had coauthored in the House. The legislation passed without a single nay vote in the chamber. However, when an appointed subcommittee “ironed out” the differences between the Senate and House versions of the bill, the Olivo amendment was removed. Although the Olivo legislation did not address many problems that Texas school children face, we believe that its passage could have constituted a first step in promoting a more just, valid, and pedagogically robust assessment system.

Reframing the Debate

An excellent point of departure for crafting a new discourse on accountability is the most high-profile, scholarly debate in Texas, appearing in the December 2000 and December 2001 issues of the *Phi Delta Kappan*. In this section, I summarize the positions and concerns expressed by Scheurich, Skrla, and Johnson (2000), the response by Valencia, Valenzuela, Sloan, and Foley (2001), and Scheurich and Skrla’s (2001) rebuttal. Revisiting this debate provides an opportunity to examine the assumptions that underlie Texas-style accountability and to contrast those with the assumptions that underlie the alternative approach of assessment based on both authentic assessment and multiple compensatory criteria. Identifying and evaluating these assumptions is more than an academic endeavor. What is meant and not meant, and what is articulated and what is left unspoken when we declare ourselves for or against “accountability” have direct bearing on real-world public debate and policy making.

In “Thinking Carefully About Equity and Accountability,” Scheurich and colleagues (2000) present a general argument justifying the Texas accountability system by pointing to the state’s “miserable record” in educating children of color, including a historic problem with deficit thinking and subtractive schooling. Moreover, they posit that with an emphasis on results, as measured by test scores, it will be possible to establish a new “national equity norm” for measuring poor and minority

childrens' achievement across the fifty states. By this they mean that equity in achievement would be a taken-for-granted assumption in the context of a national consensus to eliminate the achievement gap between Anglo children and children of color.

Scheurich and colleagues (2000) then propose that we are at a critical crossroad; that for the first time, there is an alignment of factors that make the historically elusive goal of educational equity an attainable possibility. First, the high level of public attention currently being paid to poor and minority children is unprecedented. Second, both major U.S. political parties now support a major public commitment to ensuring high academic performance from poor and minority youth. Last, in some states, including Texas, where accountability systems are in place, there are signs of rising academic achievement among some poor and minority children in particular schools and districts. As evidence of this last point, they refer primarily to their own research among Texas schools and districts populated by low-income children of color, where they have witnessed improved test performance, as well as rising test scores on the National Assessment of Educational Progress (NAEP), often referred to as the "nation's report card." Scheurich and colleagues' research findings, based on interviews with school and district administrators, reveal that it was not until the establishment of the accountability system that administrators came to believe that they could hold all children to high standards and achieve positive results.

In the remainder of the article, Scheurich, Skrla, and Johnson (2000) turn their attention to elaborating a taxonomy of multiple, interacting components of accountability, through which "equity effects" should be both studied and judged. Their taxonomy consists of (1) system elements, covering comprehensive factors that would allow equity to be judged on the basis of similar criteria across all fifty states; (2) curriculum issues, particularly the relationship of state tests to state academic standards; (3) teaching issues, including how accountability system components affect instructional quality; (4) assessment issues, especially the validity of tests that purport to measure learning; and (5) other issues, such as whether some system components are increasing dropout and pushout rates within the poor and minority segments of the student population. With regard to this last point, they cite Haney (2000), who makes the case that increasing dropout rates in Texas reflect collateral effects of high-stakes accountability systems.

This taxonomy leads Scheurich and colleagues (2000) to suggest that polarizing "either/or" stances on accountability, which they believe our position represents (particularly McNeil & Valenzuela, 2001), are ill

conceived. Empirical observations of failure in one system component (e.g., teaching to the test and narrowing curricula), they argue, should not be read as a failure of accountability as a whole, but rather as the failure of a single component.

Valencia, Valenzuela, Sloan, and Foley (2001) respond to Scheurich and colleagues (2000) by first pointing to common ground. We begin by concurring with their assertion of the state's "miserable record" on educating children of color and the deficit thinking and subtractive schooling practices that have been associated with this history. Although we support the notion of a national agenda to promote equal educational opportunity, we express concern over the use of standards-based reform (to which high-stakes tests are frequently attached) to achieve the new national equity norm envisioned by Scheurich, Skrla, and Johnson (2000).⁸

We then list what we see as the flaws in their "historical possibilities thesis." We discount both the notion of an unprecedented level of public attention and the assertion of a major public political commitment to poor and minority children. Public attention to these children has persisted over decades; and campaign rhetoric does not necessarily translate into equitable reform. We assert that test-centric systems of accountability treat only the symptoms, not the root causes of low student achievement. We also note both political parties' evident lack of will to use already proven means to enhance student performance (e.g., smaller class and school sizes, and improved teacher quality).⁹

We respond to Scheurich and colleagues' (2000) final claim of improved academic performance and substantially improved equity (i.e., reductions in the achievement gap) by noting that their own published research regarding the benefits of accountability (Skrla, Scheurich, & Johnson, 2000b) suggests that empirical examples of high-minority, high-performing schools are hard to come by, at least in Texas. After ten years of accountability and claims of success, we would expect to find a large number of districts with such schools. A search by Scheurich, Skrla, and Johnson (2000) yielded a total of 30 or (2.9 %) of Texas' 1,041 districts. Generalizing, as they do, from outlier districts like these to the state as a whole is problematic. Moreover, external data on those districts point to questionable progress. For example, they note positive outcomes in the Aldine Independent district where only 53.1% of the students have graduated. In a national study of 100 large urban districts cited by McNeil (this volume), Aldine's high school completion rate ranks 80th. Indeed, among the nation's 14 districts with the lowest graduation rates, six of these were in Texas. As of this writing, Texas' lofty claims about the

Houston Independent School District as a reformed district are being significantly challenged (Schemo, 2003a, 2003b; Winerip 2003).

With respect to the NAEP, we acknowledge improvements in scores among Texas students, but we also point out that Barton's (2001) research suggests that in 47 states, including Texas, an achievement gap persists on the NAEP between students in the top and lowest quartiles of White and minority students at the fourth-grade (mathematics and reading) and eighth-grade levels (mathematics) (also see McNeil, this volume). We respond to the allegation of polarization by reiterating that we are for accountability, while implicitly suggesting (through our alternative vision) our opposition to high-stakes testing.

We support an alternative model of accountability that is based on a framework developed by the Coalition for Authentic Reform in Education (C.A.R.E.). The model is centered around the notion of *local control over assessment* (for details, see www.fairtest.org/arn/masspage.html). Standardized examinations would be administered primarily to test for literacy and numeracy, and would be used in combination with multiple sources of information about a student's academic performance when making retention, promotion, and graduation decisions. These non-test sources would include student exhibitions, portfolios, products, and performance tasks. Under this model, state curricular standards reflect an essential but limited body of knowledge and skills that call for broadly defined competencies. Relying on a less regimented, more localized curricular focus that is authentically situated within the values and opportunity structures of communities would promote local innovation and a democratic spirit of parental and community participation, resulting in a more robust and authentic form of assessment.

This alternative proposal also calls for the establishment of quality review boards (QRB) at state and regional levels. At the state level, the QRB would bear the primary responsibility for annually evaluating the quality and availability of resources in schools and districts across the state. At the regional level, QRBs would consist of such stakeholders as teachers, parents, administrators, business representatives, and state education staff. These groups would bear the responsibility of reporting annually to their communities regarding student progress. Schools, in turn, would assume decision-making responsibility for curriculum and assessment, but these decisions would be subject to the review and approval of their districts and regional QRBs. Accountability is thus not to the state, but rather to a citizen public and the communities that schools serve.

Because of space limitations in the *Phi Delta Kappan*, current nationwide efforts toward authentic assessment were not discussed. Rhode

Island and Maine have been developing and experimenting with alternative assessment models. In Rhode Island, students entering the ninth grade in September 2004 will be subjected to alternative graduation criteria rather than to exclusive, test-based requirements (Rhode Island Board of Regents for Elementary and Secondary Education, 2003). That is, they must exhibit proficiency in a common core curriculum that includes the arts and technology. Schools have the option of using results of the state assessment as part of the graduation criteria, but they may not count for more than 10 percent of all the weighted factors contributing to the graduation determination.

Maine's assessment model is founded on overarching systemic change. By law, each district is to develop and implement a Comprehensive Local Assessment System (CLAS). The responsibility of the CLAS is to assess students' performance in relation to the content standards in eight content areas. Each CLAS was required by law to be in place by the end of 2002–2003 and it had to be certified by its superintendent as valid and reliable. Accordingly, each CLAS must guide teaching and learning, certify students' achievement in relation to the standards for high school graduation, and hold its district accountable for students' performance. In addition to administering the Maine Educational Assessment (MEA), students may be assessed with teacher-developed tests, nationally normed achievement tests, exit presentations, portfolios, district-wide administered and scored writing prompts, and other locally developed performance assessments (www.elm.maine.edu/assessments/class/). Another feature of the CLAS is that test scores are used constructively to trigger resources to schools in need.

For both Rhode Island and Maine, scoring children according to multiple assessment criteria is challenging but not impossible with computer technology. In Maine, some districts are developing their own software whereby student report cards are electronically generated. Each report card evaluates the students' mastery of state objectives according to various forms of assessment. By definition, although the model of assessment and accountability that a particular state takes would differ, the principle of multiple assessment criteria (not multiple tests) on which to evaluate students' work would be a common denominator. Moreover, such assessment alternatives are compatible with the requirements of the NCLB legislation.

In "Continuing the Conversation on Equity and Accountability: Listening Appreciatively, Responding Responsibly," Scheurich and Skrla (2001) take issue with our criticisms in several areas. First, they sense that we misunderstand their commitment to addressing systemic problems.

Accordingly, they cite a litany of historic inequities in need of redress: high teacher turnover; a lack of expertise among many teachers in high-poverty schools; chronic tracking, with children of color excluded from high-level classes; unjustifiably high rates of children of color in special education classes; low per-pupil expenditures; subtractive educational classroom experiences in which teachers fail to use the language and culture of the students; and power evasiveness among White teachers who say that they do not see color. Moreover, in light of their expressed view that results from state accountability systems do not exist in isolation from the effects created by other changes in state policies (such as increased state funding for public schools, mandatory class-size reductions, etc.), they wonder why “everyone seems to hear only that we support accountability” (p. 324).

Schuerich and Skrla also take issue with our characterizing their research results as an outlier effect. They suggest that to characterize schools and districts in this manner is to invoke a bell curve assumption, whereby most schools located in the middle of the curve are presumed to be doing little to nothing to promote student achievement. Finally, in response to the C.A.R.E. proposal, they urge us not to “romanticize” the notion of local control. They refer to the historic necessity for communities of color to seek statutory assistance from the federal government because of racist and exclusionary practices at the local level. They note that historic civil rights legislation was literally forced on local communities steeped in a culture of racism, but, they maintain, over time these changes came to be viewed positively. Top-down accountability, they suggest, should play out similarly.

Continuing the Conversation

Here, I offer some clarification of our position regarding the role of high-stakes testing in educational accountability. First, although we all share a deep concern over historic inequities, Scheurich and Skrla (2001) respond to these paradigmatically. They suggest adding opportunity-to-learn (OTL) measures to accountability scores for all schools and districts. Presumably, this would involve using an accountability reporting matrix that incorporates either additional or weighted results as part of the calibration of test scores according to available resources. This recommendation is consistent with their assumption that measurement can resolve the problem of unequal opportunities to learn the state curriculum.

Our continuing concern is that, at least in Texas, such “mathematical leveling” in performance reporting still would be premised on a misuse of tests through high-stakes testing. Scheurich and Skrla (2001) express some reservations about test validity, but generally they assume that the tests are valid—or valid enough—for accountability purposes. We disagree. A single measure of academic performance is not a valid indicator of achievement and thus cannot be an acceptable basis for holding the test taker or her teacher or his school or their district “accountable.” The test makers themselves do not suggest otherwise (Valenzuela, 2002). Moreover, the act of adjusting scores by adding or subtracting OTL units could have the insidious effect of replacing constructive action—such as increasing funding or adding staff to resource-poor schools—with complex accounting techniques. Perhaps because their view of the state’s testing system is based on interviews conducted with top-level administrators (school and district officials) rather than with teachers and parents, Scheurich and Skrla do not see Texas-style accountability as problematic. They also place great stock in the power of measurement to increase attention to existing problems and stimulate the downward flow of resources from district superintendents to classroom teachers. They anticipate that where the numbers show a need for improvement, more resources will be allocated for more hours of instruction. The result will be more equitable educational outcomes.

We recommend that instead of bringing in yet another measure of inequities as part of an accountability rubric, the state’s obligation would be to redistribute material and human resources across districts and also to assist districts in redistributing resources equitably. Overhauling enduring structural problems that are the root cause of equally enduring educational inequities is the change that we seek. Rather than continuing to rate and diagnose the inequities, we call for putting in place a more democratic and balanced model that is as concerned with inputs (resources) and processes (quality of instruction) as our colleagues are with outputs (test scores). This case is forcefully made by Ruiz de Velasco (this volume), who maintains that the basic elements for the academic success of English-language learners are simply not in place. Our aim is not to banish testing. Rather, as explained earlier, in our vision of equity and accountability, test scores would be only one of many different kinds of assessment criteria.

With regard to Schuerich, Skrla, and Johnson’s (2000) concern that readers seem to misunderstand their position, two explanations come to mind. First, in their discussion of OTL, the authors themselves suggest the centrality of accountability. A tortured circular logic guides their

thinking: If there are persistent inequities, we need to measure them to make them “visible, obvious, and public” (p. 324). If there is a reduction in inequities, it is because they were made visible, obvious, and public. Perhaps because accountability “worked” in the schools where they conducted their research, Schuerich and colleagues seem unable to set aside their silver-bullet, cause-and-effect approach long enough to consider not only that responses to test-based accountability are varied, but also that in many instances accountability actually widens the achievement gap (see McNeil, this volume).

When the authors elaborated their five system components in their first article, they failed to address whether the high-stakes tests in Texas are appropriate, either as sole or primary indicators of academic competence. Their utter silence on this issue undermines their stated agreement with scholars (e.g., Hood, 1998) who argue that the current state tests favor middle-class Whites and are culturally biased against children of color. Not addressing this aspect of high-stakes testing suggests that Schuerich and colleagues believe that these considerations are secondary to the greater goal of equity that results-based accountability, in their view, promises. Similarly, despite their expressed concern over the absence of culturally and linguistically relevant curricula in schools, the authors do not include them into the curriculum section of their taxonomy. Finally, when Scheurich and colleagues acknowledge the collateral effects of testing reported in our research (McNeil & Valenzuela 2001; McNeil, 2000), they view this more as a lapse or breakdown within a system component than as evidence of a flawed design.

Logic alone dictates that when assessment doubles as the tool for monitoring student, teacher, school, and district quality, assessment is vulnerable to corruption. A strength of the C.A.R.E. model of accountability is that it separates these functions of assessment from monitoring. Scheurich, Skrla, and Johnson’s model of accountability posits conceptually distinct components, but then negates this separation by failing to make provisions for the varying influence of the determinative weight of high-stakes tests. For example, curriculum issues (component 2), such as teaching to the test and “dumbing down” course materials, are intimately related to assessment issues (component 4) in a context in which high-stakes consequences for both children and schools are attached to children’s scores.

For those of us who share a deep concern over extant inequities, the fact that Scheurich and colleagues’ framework stops just short of addressing the adverse impact of high-stakes testing on minorities provokes an especially strong reaction to their thesis of historic possibilities. This

aspect of their position also encourages our reading of their model of accountability as uncritically wedded to the notion of high-stakes testing. Such facile coupling of high-stakes testing to accountability presumes, on the one hand, that this kind of testing is a necessary feature of accountability systems and, on the other hand, that such tests actually measure what they purport to measure.

Our proposal of a more democratic form of accountability separates the concept of high-stakes testing from accountability. We posit that it is possible to have an accountability system that does not harm children. Stated more positively, we insist that accountability can and should be consistent with the use of locally derived, authentic assessments, academic rigor, and rich learning experiences. Children need not bear the brunt of unequal OTL through their graduation or nongraduation, or their promotion or nonpromotion. Absent these considerations, Scheurich and colleagues' vision of equity and accountability rings hollow.

The issue of outlier effects needs clarification. We did not intend to suggest the existence of a bell curve along which schools and districts could be ranked. Rather, our concern was and is with the generalizability of the claim that accountability works to change mindsets that previously were deficit oriented and subtractive. Does this mean that in all other schools where performance ratings have shown improvement, changed perspectives coupled with a marshaling of resources have been the decisive factor? Does a marshaling of resources signify a fundamental transformation of a poorly performing school into one that is staffed with certified and trained teachers, more advanced curricular offerings, well-stocked libraries, and fully functioning science laboratories? Or does it mean the channeling of dollars into the purchasing of test prep materials and activities of limited instructional value (see the chapters by McNeil; Hampton; Alamillo, Palmer, Viramontes, & García; and Sloan)? Does a marshaling of resources for targeted improvement qualify such schools and districts as nonracist? As culturally and linguistically relevant? What are we to make of the fact that the state-mandated, culturally chauvinistic curriculum remains intact?

Another concern we registered was that ten years of accountability in Texas have done little to alter the state's educational landscape. When Scheurich, Skrla, and Johnson's (2000) search for high-performing schools and districts attended by children of color, many of them poor, yields a subset of fewer than 3 percent of all Texas schools, this renders untenable their claim that accountability results in increased equity. In a state as large as Texas, we are not surprised that there would be entire

districts where accountability data have been used responsibly. However, that some schools and districts are acting responsibly does not allay our concerns over test abuse, collateral effects, or adverse impact on poor and minority youth.

Finally, in cautioning us not to romanticize the notion of local control, Scheurich and colleagues cite top-down reforms accomplished by the civil rights movement. They note that people in power complained about integration and Title IX but grew accustomed to these changes over time. This analogy implies that the “movement” toward standards-based reform emanates from the mobilization efforts of poor, disenfranchised communities. This characterization could not be further from the truth.

The standardized accountability system has been unilaterally and undemocratically foisted on poor and minority communities. Poor, minority and immigrant communities may have been led by the rhetoric of “standards-based reforms” to believe that the state’s accountability system would deliver to their communities the educational resources to help their schools—and their children—receive a highly academic education, one that attained high academic standards more typical of the education of children in more privileged districts. Little prepared, many of these parents felt encouraged to support the standardization of their children’s schools and the reduction of their children’s educational achievement to a single score on a single measure. A single-measure accountability system based on a standardized test, however, occupies no organic space in the wishes, desires, or mobilization efforts of poor, minority communities in Texas or elsewhere (see especially McNeil, 2000; Trujillo, 1998).

The case of desegregation is instructive, though, to the extent that it reminds us that when the statutory engines of change commanded restraint, they targeted not the victims of Jim Crow, but rather the perpetrators of discriminatory practices. Just as it would have been unreasonable—indeed, nonsensical—for either the lowest-level workers or business clientele to have assumed the primary burden of change to integrate lunch counters, hotels, and other public establishments, so too is it unreasonable for both children and their teachers to bear the burden of change under the banner of accountability. Using a market metaphor, it is similarly illogical to make either students or teachers responsible for the quality of their product when they do not control the resources or flow of finances to which the outcomes are tied (Dye, 2002).

It is important to note, however, that from a rational-choice, decision-making perspective (which is the underlying framework of the voucher/school choice movement), holding children—and by extension, their families—responsible for test score “products” does make sense.

That is, families *as consumers* can make use of the test scores and school ratings that accountability systems provide as a basis for their decisions regarding which schools their children should attend. From this perspective, school quality is less an issue of structural inequities, and more a matter of individual choice based on the “objective” information that accountability provides. Hence, an unintended consequence of the accountability model advanced by Scheurich, Skrla, and Johnson (2000) is a latent effect of advancing a conservative, consumer-oriented agenda that seeks nothing less than the privatization of public education (Labaree, 1997; Meier, 1995; Valenzuela, this volume). In light of their expressed concern for minorities, it is ironic, then, that Scheurich and colleagues’ model should open the door to consumer-minded, often deficit-oriented individuals who have abandoned even the pretense of striving for equity.

One final point: Rather than being “romantic,” the C.A.R.E. proposal implicitly reflects a studied consideration of the appropriate role of the state. Its aim—to limit the power of the state to make assessment and curricular decisions—is hardly an endorsement of a conservative, states’ rights or return-to-local-control perspective. The equitable distribution of resources is an appropriate role of the state. The premise underlying our position is that the community is in a better position than the state to define not only what constitutes a quality education but also to assess children’s talents, abilities, and potential. In contrast, one-size-fits-all performance standards conceived and administered at the highest bureaucratic levels encourage a regimented, test-driven curriculum and increase the chances that inaccurate decisions will be made on children’s behalf.

Guiding Assumptions

The preceding analysis suggests, first, that proponents of Texas-style accountability view high-stakes testing as the driver for equity. Second, it suggests that they conceptually equate high-stakes testing with educational accountability, so much so that our anti-high-stakes-testing position gets construed as anti-accountability. This lack of precision helps explain why we are often accused of polarizing the debate on accountability. The truth of the matter is that although each side in this debate uses the same or similar terminology, we often mean different things. Third, for all their misgivings concerning the testing system’s imperfections, they view “the test” as both an essential feature of accountability and a valid measure of educational knowledge and skills. Fourth, alternative epistemologies, values,

and ways of knowing are secondary to the greater goal of equity that results-based accountability promises. And last, control over assessment properly resides in the hands of the state.

Opponents of Texas-style accountability bring a different set of assumptions to the table. First, we view high-stakes testing as unethical and inappropriate for measuring students' talents, abilities, and potential. Moreover, through collateral effects (such as narrowing curricula and marginalizing students), Texas-style accountability systems rob children of a quality education, fostering their psychic, emotional, and, sometimes, literal physical withdrawal from the schooling process.

Second, we distinguish accountability from high-stakes testing. Our insistence that accountability can be pursued without making children bear the brunt of it through their retention/non-retention or graduation/nongraduation on the basis of a single exam may be our most important contribution to the debate. Getting this message across to policy makers and the lay public is crucial because, at least in Texas, these stakeholders often equate accountability with high-stakes testing.

Third, using a single test to judge whether a student has mastered the state curriculum is inherently invalid. More valid decisions are made when educators use multiple and authentic sources of assessment of student learning. Additionally, coupling the assessment instrument with the monitoring function of accountability not only corrupts the assessment, but also makes children assume primary responsibility for both educational quality and inequality.

Fourth, Texas-style accountability is subtractive to the degree that the curriculum and the testing instrument reinscribe relations of domination and subordination between Anglos and Latina/os through a culturally and linguistically chauvinistic curriculum that privileges the English language while devaluing fully-vested bilingualism and biculturalism. In the current incarnation of accountability, alternative epistemologies or ways of knowing are not encouraged because the system is said to be culture-blind. Last, control over assessment properly resides not at the end of the long arm of the state, but rather in the caring hands of those who are in the best position to know our children, namely, their teachers, parents, and community. With our proposal for a more reciprocal and responsible form of accountability, the state's proper role is recast in terms of its central obligation, which is to provide the material and human resources that schools need to promote their equity goals.

The structuring out of culturally relevant schooling through the standardized educational experience that is reflected in Texas' current model of accountability results in enormous sacrifices that we as a society

cannot afford (see McNeil, this volume). These include closing off teaching that draws from the richness of children's experiences, depriving them of our deepest and broadest understandings of our particular and shared cultural heritages in an increasingly interconnected global economy and planet. Standardized curricula, as argued in the chapters that follow, also lessen children's access to our greatest scientific achievements and our myriad ways of knowing (see Padilla, this volume).

Although the C.A.R.E. proposal insufficiently addresses the importance of culture in education, it creates space for this by calling for other ways of knowing, and teaching, to children's abilities and potential. Particularly in a context of rapidly changing demographics where the increasing presence of Latinos is reshaping entire communities, our sense of what children can learn should be based less on the limited goal of whether they can pass centrally designed, computer-scored tests, and more on a rigorous and challenging vision of schooling. Neither children nor their communities, languages, or identities represent liabilities, or "problems" to be overcome. From an asset-based perspective, they offer needed qualities in an increasingly globalized world.

Within an additive schooling model, caring for children in an authentic manner means honoring their community-based identities in a respectful, relational manner—*con propio respeto* (with due respect). Through a culture of engagement as Padilla (this volume) suggests, we can rekindle our democratic impulse to both search out our commonalities while respecting difference and affirming as Americans our commitment to shared principles like hard work, fair play, tolerance, due process, and social and economic justice. Resistance to high-stakes testing is thus not a ploy to water down quality or requirements, but rather a call to citizens and parents to exert democratic and community-based authority to rectify a system out of balance.

CHAPTER SUMMARIES

In chapter 2, "Performance-Based School Reforms and the Federal Role in Helping Schools That Serve Language-Minority Students," Jorge Ruiz de Velasco assesses the federal role by focusing attention on perhaps one of the most vulnerable student populations in our schools, namely, immigrant, English-language learners (ELLs). Ruiz de Velasco makes the case that typical assumptions that hold for English-speaking children's classrooms do not generally hold true for those filled with ELL youth. He investigates assumptions undergirding standards-based reforms and

suggests reasons why those reforms that rely on high-stakes testing are not likely to be sufficient in meeting the needs of ELL youth. For starters, ELLs vary widely with respect to the number of languages that they speak, their prior levels of former schooling, their parents' levels of education, and the kinds of literacy practices that take place in their homes. Indeed, many enter our schools significantly under-schooled. Such students not only find it difficult to work at age-appropriate levels in required subjects, they also have difficulties doing so in their native languages. Nor are schools sufficiently equipped with trained staff and appropriate technologies for teaching to this wide diversity. Ruiz de Velasco also elaborates on what he thinks the appropriate role of the federal government should be at this historic juncture.

Chapter 3, "Faking Equity: High-Stakes Testing and the Education of Latino Youth," by Linda McSpadden McNeil, carries forward the critique of an allegedly objective accountability system by providing a thorough analysis of numbers-based accountability in Texas. She finds that accountability not only fails to tell the whole story, but depends on a partial accounting for the story to be true. Specifically, McNeil challenges the claim widely made in Texas that achievement is improving. She finds that on every indicator of academic achievement other than the TAAS, school children in the state not only register poor performance levels, they also are growing increasingly weaker academically.

McNeil portrays accountability as a system of controls governing nearly every aspect of public schooling in Texas. Moreover, these controls hinge on a standardized test that all children must take, which then becomes a basis for rating schools. This results in pressures to raise scores by narrowing curricula and marginalizing students who threaten school ratings. She maintains that the state is faking its claims of improving test scores. McNeil counters with evidence from other indicators, including SAT, ACT, and NAEP scores, and high school graduation rates. She concludes by suggesting that the evaluation of adults' (district administrators, school principals, and classroom teachers) performance must be decoupled from children's test scores. She also calls for an *educación* model of schooling that respects the cultures and values of children's families. In short, McNeil makes a persuasive case for why Texas-style accountability is not a model that other states should follow.

In chapter 4, "Texas' Second Wave of High-Stakes Testing: Anti-Social Promotion Legislation, Grade Retention, and Adverse Impact on Minorities," Richard R. Valencia and Bruno J. Villarreal review more than five decades of research on grade retention. They find clear evidence of its harmful impact on students, particularly with respect to academic

outcomes. They point out that results in social science are rarely unequivocal. Yet study after study shows that although a few students perform better after being retained, most demonstrate either no progress or lower academic achievement when followed over time. Indeed, grade retention significantly increases the likelihood that students will drop out.

Valencia and Villarreal's analysis of projected third-grade retentions in Texas as a result of the state's new anti-social promotion policy provides sobering evidence of the disparate impact of SB 4 on Latinos, particularly Mexican Americans, and African Americans, relative to Anglos. They critique policy makers' penchant for attributing minorities' chronically low achievement to individual-level factors such as cognitive ability, motivation, and family background characteristics. Inasmuch as minority youth attend poor, segregated schools staffed by teachers who frequently are not certified to teach the courses they are assigned, high failure rates on standardized tests are scarcely reducible to individual aptitude, motivation, and decision making. Instead, poor test performance, alongside retention and dropping out, are better construed as symptomatic of many children's inferior schooling experiences (see also Valencia & Bernal, 2000).

In chapter 5, "Playing to the Logic of the Texas Accountability System: How Focusing on 'Ratings'—Not Children—Undermines Quality and Equity," Kris Sloan reports on his investigation of a Houston-area elementary school and school district. In considering accountability proponents' claim that it can be used to leverage educational reform, Sloan cites examples of places where such leveraging has occurred, drawn from research by Scheurich, Skrla, and Johnson (2000). He contrasts these findings with those of McNeil and Valenzuela (2001), whose investigation of accountability shows the system contributing to a reduction in the quality and quantity of curricula delivered to poor, minority children. Sloan reconciles these two strands of research by noting important differences in the focus of the studies. Scheurich and colleagues' research is top-down, gathered mostly from district-level personnel, while McNeil and Valenzuela provide a bottom-up view from the classroom. Sloan maintains that these divergent approaches help explain divergent conclusions. He situates his own two-year study within the latter approach.

As a matter of interest, if Sloan had conducted only a one-year ethnographic study, he would have concurred with Scheurich and colleagues' position. During the first year of his two-year study, teachers at Glendale Elementary school used the accountability system to successfully leverage reform through authentic approaches to curricula, activity-based learning, and an ethic of care. During the second year of the study,

however, their reformist agenda was subverted into a ratings-focused response, largely through district-level mandates. Higher test scores and an accompanying positive school rating were achieved, but the cost was high—Glendale was transformed into a “factory model” school that aligned “people, customers, strategy, and processes” (Sloan, 2002, p. 171). The earlier, child-centered reform efforts were abandoned. The “TAAS discourse” that ensued, Sloan maintains, was able to dominate issues of educational quality precisely because of the composition of the school’s student population—mainly poor children of color.

In chapter 6, “Standardized or Sterilized? Differing Perspectives on the Effects of High-Stakes Testing in West Texas,” Elaine Hampton presents interview, survey, and focus group data from nine elementary and three middle schools in the Ysleta Independent School District in El Paso. Hampton’s survey respondents and focus group members reported their perceptions of the effects of mandated testing on curriculum, students, and school climate. The data provide support for Sloan’s view that perceptions of accountability are affected by individuals’ position in the system. Across the schools that Hampton studied, administrators were much more positive about the effects of mandated testing than were the teachers in their schools.

Hampton also builds on Sloan’s and McNeil’s contributions with her findings that in response to administrators’ demands, a test-centric curriculum unfolded at the classroom level. In one middle school, for example, teachers from every class in all subject areas taught math during their own class time. Students also took one to three math classes, per week (during their two-day block schedule), that focused exclusively on TAAS objectives. Professional development for teachers in the middle school was reduced to establishing teaching teams whose members then trained colleagues to design TAAS-like tests for use in their own classrooms. The elementary schools in Hampton’s sample followed a similar path, supporting her contention that the accountability system is creating a standardized, sterilized curriculum and classrooms where rote, fragmented learning and monotony prevailed.

Laura Alamillo, Deborah Palmer, Celia Viramontes, and Eugene E. García highlight in chapter 7, “California’s English-Only Policies: An Analysis of Initial Effects,” policy and practice changes that have occurred in bilingual education programs under the rule of California’s Proposition 227. Passed by voters in 1998, Proposition 227 restricts the use of the student’s primary language in classroom instruction. Alamillo et al. collected interview data from teachers and principals statewide who work with language minority, elementary school students. Some of the

educators in the sample continue to teach in their students' primary language because they work at sites where exemptions have preserved the bilingual programs, despite the recent policy mandates. Others work in schools that have dismantled their bilingual education programs. Across these two settings, increased frequency of assessments in English have resulted in a push for more English curricula and instruction, as well as in greater pressures on teachers to transition students into mainstream English to raise test scores.

Alamillo and colleagues find, as do McNeil, Sloan, and Hampton, that these pressures have reduced both the quality of learning and teacher autonomy, while increasing control over teacher practices. Teachers in schools with waivers that allow them to continue teaching bilingually even in the wake of Proposition 227 felt that the state's English-only testing system "was even more damaging to their primary language programs than Proposition 227 had been" (p. 217). Several participants in the study also cited racism against Latinos as an explanation for increased mass testing in English only.

In chapter 8, "The Centurion: Standards and High-Stakes Testing as Gatekeepers for Bilingual Teacher Candidates in the New Century," Belinda Bustos Flores and Ellen Riojas Clark examine the records of 20 students from a state university teacher education program who had been teaching in public schools located in a large, urban district. The analysis reveals how accountability logic attaches to teacher certification requirements to the detriment of prospective Latino, bilingual college students in pursuit of bilingual certification. Bustos Flores and Riojas Clark show how these students got caught in a web of university policies that were not designed for teachers who were seeking bilingual certification.

Specifically, by imposing a qualifying test, the university screened out potential failures on the "real" (state-mandated) exit test upon which both program accreditation and federal aid are based. This (along with other test-centric bureaucratic hurdles) negatively impacted Latina/o teachers seeking bilingual certification because it required them to take an exam outside of their specialty area; the exam presumed training commonly given to mainstream prospective teachers but not typically included in bilingual teacher programs. The test did not reflect, for example, differences in Spanish language arts training compared to English language arts training.

As a consequence of this and other gatekeeping mechanisms, many of these demonstrably qualified preservice teachers were unable to find stable employment. No matter how effective they were in the classroom, they had to take and pass the state-mandated exit exam to remain in their

teaching positions. Bustos Flores and Riojas Clark note that at a time when there is a critical shortage of Latina/o teachers, these roadblocks preserve teacher training programs in the state universities but disproportionately sacrifice Latina/o students in the process.

The following chapter, “‘High-Stakes Testing’ and ‘Educational Accountability’ as Social Constructions Across Cultures,” by Raymond V. Padilla, frames the contributions to this volume with a Latina/o epistemological perspective. He develops three major topics: (1) testing in relation to our present culture of measurement; (2) the underlying reductionism of accountability; and (3) alternative social constructions of educational reform that provide a richer sociocultural learning context for culturally diverse students. Padilla draws on the Spanish language terms *saber* and *conocer* to differentiate between objectivist and relational ways of knowing, respectively. He then shows how these epistemological differences are associated with distinctive cultural frameworks, which he terms the *culture of measurement* and the *culture of engagement*, respectively. He writes critically of the culture of measurement, which includes the currently popular model of de-contextualized accountability with high-stakes testing. The culture of measurement arises from a context-free perspective that privileges a numbers-based, objectivist way of knowing, and holds students solely responsible for school performance. It departs dramatically from the context-based culture of engagement that ascribes value and meaning to relationships and community, alongside a responsibility to care for children.

For Latina/os and other culturally diverse students, the culture of measurement can have harsh implications. An objectivist, decontextualized curriculum promotes a systemic disregard for the languages, cultures, and assets that children bring with them to the classroom, with the result that these students often are marginalized and alienated, to their academic and personal detriment. Padilla also notes the shared meaning of accountability within English and Spanish (*contabilidad*), where “to count” (*contar*) and “to recount,” either with numbers or through narrative, are alternating options. Due to the culture of measurement, however, the semantic space accorded to “accountability” in the English language has narrowed. Instead of telling a story or providing a full accounting, the meaning of accountability is now reduced to numbers-based measurement.

In the final chapter, “Accountability and the Privatization Agenda,” I make the case that the current accountability system serves as handmaiden to the privatization agenda in Texas. The political agenda behind the current accountability system became transparent while witnessing

the strong push to privatize public education in Texas through various proposals during the 2003 legislative session. I document the presence of a powerful, neoliberal political and economic elite that helped develop and now works within the current framework on educational accountability, while simultaneously supporting privatization efforts, including school vouchers.

I expound on the debate over vouchers both because of my personal involvement in this issue and because of the influential, if unfortunate, role that some Latinos from the League of United Latin Americans Citizens (LULAC), in particular, have played in this evolving debate. The narrative shows that the Latino community is positioned on the sidelines of a much larger drama characterized by a right-wing, conservative agenda that both exploits the long-standing grievances that poor and minority communities hold toward schools and districts and deploys raw, political power to accomplish its goals. This account illustrates how localized issues, people, and cultures intersect when the forces for privatization emerge in a state or community. Building on the preceding chapters, this final piece demonstrates how accountability is more about the politics of control over public education than it is about children's learning and well-being.

In conclusion, as states elaborate their testing systems mandated by the No Child Left Behind Act, we hope educators and policy makers will heed the lessons provided by Texas-style accountability. The particular legislative issues in Texas reveal the complexity of the struggle and the difficulty in changing these top-down, state-mandated systems once they are in place. Increasing numbers of parents, educators, and government officials across the United States realize that accountability systems based on standardized tests carry high costs to districts, and harmful effects on children. Also, efforts involving revision or resistance are taking many forms. In Texas, it is important to note that the locus for change has been in the legislature, where important bills originate.

The evidence and arguments in this volume make clear that to the degree that states attach high-stakes consequences to tests for children, schools, and districts, they are likely to reproduce the uneven, unjust, and unnecessary educational outcomes we have documented in Texas. For the sake of all of our children, we hope that those who design the new systems will seize the opportunity to reconceptualize accountability by first divesting it of its high-stakes testing component and developing in its place multiple and authentic sources of assessment that will provide a more valid, just, and humane picture of student achievement. The purpose of accountability is to make sure that the educational system serves

the children. Our earlier statement bears repeating: Accountability is thus not to the state, but rather to a citizen public and the communities that schools serve.

NOTES

1. I use the terms Latina and Latino (or Latina/o) to be inclusive of all Latina/o subgroups in Texas. According to the 2000 U.S. census, U.S.-born Mexican Americans and immigrant Mexicans combined are the largest Latina/o subgroup, numbering 5,071,963 out of the total state population of 20,851,820 (24.3%). (Source: U.S. Census Bureau, Census 2000. Table DP 1. Profile of General Demographic Characteristics: 2000, Geographic Area: Texas.)

2. We presented this position at a Capitol Hill briefing to address the effects of high-stakes testing and social promotion on Latino youth on Friday, March 15, 2002. The event was conducted in conjunction with the Congressional Hispanic Caucus, the Hispanic Education Coalition, and the Center for Mexican American Studies (CMAS) at the University of Texas at Austin.

3. Our position on high-stakes testing is shared with all of the leading national educational and measurement organizations, including the American Educational Research Association, National Academy of Sciences, Association for Supervision and Curriculum Development, American Psychological Association, National Council of Teachers of English, National Council of Teachers of Mathematics, and the National Council on Measurement in Education.

4. I quote here from Kauffman (1999), "In 1997 at the end of the twelfth grade, approximately ten thousand seniors were still taking the TAAS Exit Test (Fassold report, PX 26). Of these, 87% were either Hispanic or African American. On this 'final' administration of the TAAS Exit Test, 41% of Whites, 32% of African Americans and 27% of Hispanics passed the test (Fassold test; Fassold report, PX 26 at pp. 8-9)."

5. Historically, the exit-level TAAS test has been taken beginning during the students' sophomore year, giving them up to eight additional opportunities to take the exam. Beginning this year, the exit exam is to be given in the eleventh grade, giving students fewer opportunities to take the state exam (see McNeil).

6. For an excellent review of this case from a research-based perspective, see the special issue of the *Hispanic Journal of Behavioral Sciences* edited by Valencia and Bernal, 2000.

7. The Urban School Alliance, a coalition of urban district administrators representing a quarter of all children enrolled in the state, made use of a clause in SB 4 that instructs the commissioner to certify whether sufficient funds have been appropriated statewide for the legislation (TEC §28.0211 [m]). The Alliance attempted to make the case that funding was insufficient. However, their concern was subsequently nullified by the commissioner who indicated that sufficient funds were available.

8. According to Ingersoll (2003), as of 2002, nineteen states have high-stakes graduation tests in place. Five states, including Texas, make grade promotion contingent on passing a test. At the school level, all schools that are now receiving Title I monies are subject to high-stakes consequences if adequate yearly progress is not achieved (see the NCLB website at <http://www.ed.gov/legislation/FedRegister/finrule/2002-3/070502a.html> for Title I regulations).

9. See Ingersoll (2003) for a national-level analysis of the determining influence of teacher quality on improved student performance.

REFERENCES

- Barton, P. (2001). *Raising achievement and reducing gaps: Reporting progress toward goals for academic achievement. A report to the National Education Goals Panel*. Washington, D.C.: National Education Goals Panel.
- Blackwell, K. (2003, March 19) Students rise to challenge of tougher test in reading. *Austin American-Statesman*, pp. B1, B6.
- Darling-Hammond, L., Aness, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teachers College Press.
- Dye, J. (2002). *A natural yet overlooked feature of accountability in public education: Accountability is supposed to go both ways*. Unpublished manuscript.
- Fassold, M. (1999). Expert report supplemental. Submitted to plaintiffs' counsel, Albert H. Kauffman. *GI Forum et al. v. Texas Educational Agency et al.* 87 F. Supp.2d 667 (W.D. Tex. 2000).
- Fassold, M. (2000). Disparate impact analyses of TAAS scores and school quality. *Hispanic Journal of the Behavioral Sciences*, 22(4), 460–480.
- GI Forum et al. v. Texas Educational Agency et al.*, 87 F. Supp.2d 667 (W.D. Tex. 2000).
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41), 1. Retrieved August 24, 2000. Available at: <http://epaa.asu.edu/v8n41/>
- Haney, W. (2001). *Revisiting the myth of the Texas miracle in education: Lessons about dropout research and dropout prevention*. Paper presented at the Dropout Research: Accurate Counts and Positive Interventions Conference, Harvard Civil Rights Project, Cambridge, MA, January 13, 2001.
- Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kluwer Academic Publishers.
- Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Committee on Appropriate Test Use, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

- Hood, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *Journal of Negro Education*, 67, 187–96.
- Ingersoll, R. M. (2003, January 9). Special report: Quality counts 2003: “If I can’t learn from you.” *Education Week*.
- Kauffman, A. (1999). Post-trial brief. In the United States District Court for the Western District of Texas, San Antonio Division.
- Labaree, D. F. (1997). *How to succeed in school without really learning: The credentials race in American education*. New Haven: Yale University Press.
- Linn, R. L. (1982). Ability testing: Individual differences, prediction and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses consequences and controversies, Part II* (pp. 335–388). Washington, D.C.: National Academy Press.
- McNeil, L. (2000). *Contradictions of reform: Educational costs of standardized testing*. New York: Routledge.
- McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers? Inequality and high stakes testing in public education* (pp. 127–150). New York: Century Foundation.
- Meier, D. (1995). *The power of their ideas*. Boston: Beacon Press.
- Meier, D. (2000). *Will standards save public education?* Boston: Beacon Press.
- Meier, D. (2002). *In schools we trust: Creating communities of learning in an era of testing and standardization*. Boston: Beacon Press.
- Murdock, S., Hoque, M. N., Michael, M., White, S., & Pecotte, B. (1997). *The Texas challenge: Population change and the future of Texas*. College Station: Texas A&M University Press.
- Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.
- Rating TAKS: State must not waver in reaching accountability goals. (2003, April 18). *Houston Chronicle*, p. 36A.
- Rhode Island Board of Regents for Elementary and Secondary Education. (2003). *Regulations of the Board of Regents for elementary and secondary education regarding public high schools and ensuring literacy for students entering high school*. Cranston: Rhode Island School-to-Career. Retrieved March 31, 2004, from <http://www.riste.org/hsrules.htm>
- Saucedo, L. M. (2000). The legal issues surrounding the TAAS case. *Hispanic Journal of Behavioral Sciences*, 22(4), 411–422.
- Schemo, D. J. (2003a). Questions on data cloud luster of Houston schools. *New York Times*. Retrieved July 11, 2003, from <http://www.nytimes.com>
- Schemo, D. J. (2003b). For Houston schools, college claims exceed reality. *New York Times*. Retrieved August 28, 2003, from <http://www.nytimes.com>
- Scheurich, J., & Skrla, L. (2001). Continuing the conversation on equity and accountability: Listening appreciatively, responding responsibly. *Phi Delta Kappan*, 83, 322–326.

- Scheurich, J., Skrla, L., & Johnson, J. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan*, 82(4), 293–299.
- Shepard, L. A., & Smith, M. L. (1989). *Flunking grades: Research and policies on retention*. London: Falmer Press.
- Skrla, L., Scheurich, J., & Johnson, J. (2000a). *Accountability for equity: Can state policy leverage social justice?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 24–28, 2000.
- Skrla, L., Scheurich, J., & Johnson, J. (2000b). *Equity-driven achievement-focused school districts: A report on systemic school success in four Texas school districts serving diverse student populations*. Austin, TX: Charles A. Dana Center.
- Sloan, K. (2002). *A ratings-focused response to the Texas accountability system and the professional lives of teachers: An ethnography*. Doctoral Dissertation: University of Texas at Austin, Texas.
- Texas Education Agency. (1996). *The development of accountability systems nationwide and in Texas (Statewide Texas Educational Progress Study (STEPS) 1)*. Austin, TX: Texas Education Agency.
- Texas Education Agency. (2003a, March 18). *Texas third-grade students exceed predictions; 89 percent pass reading exam*. Press Release. Retrieved July 14, 2003, from <http://www.tea.state.tx.us/press/89percent.html>
- Texas Education Agency. (2003b, July 23). *Passing rate hits 96 percent on crucial third-grade TAKS reading test; retention may be only slightly higher than past years*. Press Release. Retrieved August 20, 2003, from <http://www.tea.state.tx.us/press/taks31.html>
- Texas State Board of Education. (2002). *Texas Assessment of Knowledge and Skills [TAKS] standard setting: Summary of projected impact*. Unpublished document.
- Trujillo, A. (1998). *Chicano empowerment and bilingual education: Movimiento politics in Crystal City, Texas*. New York: Garland Publishing.
- Valencia, R. R., & Bernal, E. M. (Eds.). (2000). Special Issue. The Texas Assessment of Academic Skills (TAAS) Case: Perspectives of plaintiffs' experts. *Hispanic Journal of the Behavioral Sciences*, 22(4), 403–561, 423–444.
- Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage.
- Valencia, R. R., Valenzuela, A., Sloan, K., & Foley, D. (2001). Let's treat the cause, not the symptoms: Equity and accountability in Texas revisited. *Phi Delta Kappan*, 83(4), 318–326.
- Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring*. Albany: State University of New York Press.
- Valenzuela, A. (2000). The significance of the TAAS test for Mexican immigrant and Mexican American adolescents: A case study. *Hispanic Journal of the Behavioral Sciences*, 22(4), 524–539.

- Valenzuela, A. (2002). High-stakes testing and U.S.-Mexican youth in Texas: The case for multiple compensatory criteria in assessment. *Harvard Journal of Hispanic Policy*, 14, 97–116.
- Valenzuela, A., and Maxcy, B. (in press). Limited English proficient youth and accountability: All children (who are tested) count. In J. Ruiz deVelasco & M. Fix (Eds.), *School reform, immigrants, and native minorities*. Washington, D.C.: Urban Institute.
- Willingham, W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.
- Winerip, M. (2003). The “zero dropout” miracle: Alas! Alack! A Texas tall tale. *New York Times*. Retrieved August 13, 2003, from <http://www.nytimes.com>