

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 14 Number 23

September 14, 2006

ISSN 1068-2341

NCLB: Local Implementation and Impact in Southwest Washington State

Linda Mabry and Jason Margolis
Washington State University Vancouver

Citation: Mabry, L., & Margolis, J. (2006). NCLB: Local implementation and impact in southwest Washington state. *Education Policy Analysis Archives*, 14(23). Retrieved [date] from <http://epaa.asu.edu/epaa/v14n23/>.

Abstract

The research reported here is from the first two years of an ongoing and largely qualitative study to examine the impact of the No Child Left Behind federal education policy on educational practice and climate in elementary schools in two districts in southwest Washington. Based on systematic drop-in observations in classrooms and interviews with teachers and school and district administrators, data indicated that the policy had partially yielded the intended standards-based reforms but at considerable local cost. While most participating administrators described efforts to use NCLB to leverage needed change, most teachers described struggles to sustain best practice and to avoid some negative consequences to their students and schools. Administrators anticipated that resistant teachers would be nudged from the profession, and the greatest attrition among participating teachers was from the fourth-grade level at which the state's standards-based test was administered. Fourth-grade teachers particularly expressed concern about test-related stress and test-driven curricula interfering with children's individual needs and with their own ability to provide developmentally appropriate instruction adapted for their particular students. The validity and utility of test results was a local issue.

Keywords: accountability; NCLB; policy impact; school reform; teacher change.



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Mary Lou Fulton College of Education at Arizona State University and the College of Education at University of South Florida. Articles are indexed by H.W. Wilson & Co. Please contribute commentary at <http://epaa.info/wordpress/> and send errata notes to Sherman Dorn (epaa-editor@shermamdorn.com).

“Sin abandonar ningún niño:” Implementación e impacto locales en el sudoeste del Estado de Washington.

Resumen

Esta investigación presenta datos parciales obtenidos durante los primeros dos años de un estudio en curso y mayormente cualitativo con el objetivo de examinar el impacto de la política federal de educación “Sin abandonar ningún niño” (NCLB) sobre la práctica educativa y el clima en las escuelas primarias de dos distritos escolares en el sudoeste del estado de Washington. A través de visitas y observaciones sistemáticas en salas de clase y entrevistas con docentes y administradores de las escuelas y del distrito escolar, los datos obtenidos muestran que la política federal de estándares NCLB produjo las reformas previstas pero a un costo local considerable. Mientras que la mayoría de los administradores escolares que participaron describieron los esfuerzos de utilizar NCLB para estimular cambios considerados necesarios, la mayoría de los profesores describieron problemas para sostener las “buenas prácticas” recomendadas y evitar algunas consecuencias negativas para sus estudiantes y escuelas. Los administradores anticiparon que los profesores que opondrían mayor resistencia serían gradualmente desplazados de la profesión, y que el agotamiento más grande se daría entre los docentes de cuarto-grado ya que es en ese grado donde se administra los exámenes estatales estandarizados de NCLB. Los docentes de cuarto-grado expresaron preocupación por la tensión nerviosa que generaba los exámenes y los planes de estudios organizados de acuerdo a los exámenes estandarizados, que además interferían con las necesidades individuales de los estudiantes y con la capacidad de los docentes de proporcionar una instrucción apropiada al desarrollo y adaptada a las particularidades de sus estudiantes. La validez y utilidad local de los exámenes estatales estandarizados fue considerada un problema.

Having a country this size with this degree of diversity and saying, “By golly, everybody will meet a standard,” is a huge thing. It’s wonderful, but it’s huge. It falls to a lot of big and little districts to somehow make that happen. (Highland district administrator, June 23, 2005)

The most forceful federal education policy in forty years, the No Child Left Behind statute (NCLB, 2001) mandates universal literacy and numeracy by 2014 at third through eighth grade levels as measured by standards-based state tests,¹ the scores calibrated against scores on the National Assessment of Educational Progress (NAEP) for credibility purposes. Determination of the effectiveness of the federal policy requires local investigation because the primary intended effects are local: the refocusing of educational delivery in schools and classrooms to ensure attention to all students, including the historically underserved, and the raising of individual student achievement. Beginning one year after NCLB’s implementation, the follow-along policy study reported here,

¹ All states except Iowa include standards-based testing as primary components of their educational accountability policies, although many of these 49 states administer norm-referenced rather than standards-based tests at some grade levels.

undertaken in two school districts in southwest Washington state, documents NCLB's local implementation and impact.

In the past quarter-century, studies of educational reform initiatives have documented their increasing politicization (e.g., Cook, Fitton, Viator, Phillips, & Gertz, 2001; Cuban, 1988; Fuhrman & Elmore, 1990; Wise, 1979) and many disappointments (e.g., Booher-Jennings, 2005; Goodson & Foote, 2001; Lee & Wong, 2004; Stringfield & Yakimowski-Srebniak, 2005). The exclusion of educators from higher level policy-making has been shown to result in overestimations of local capacity (e.g., Borko, Wolf, Simone, & Uchiyama, 2003), conflicts with local policy (e.g., Shipp, 2003, Tyack & Cuban, 1995), local unwillingness to implement external initiatives (e.g., Darling-Hammond, 1990; Fullan, 1991; Sarason, 1982, 1990; Sipple, Killeen, & Monk, 2004), and attempts by local administrators to buffer classrooms from anticipated consequences (e.g., Rossman, Corbett & Dawson, 1986). District- and school-level implementation has often taken the form of superficial compliance (e.g., Burlingame, 1993; Raywid, 1990), uncomfortable conglomerations of old and new practices (e.g., Louis, Febey, & Schroeder, 2005; Tyack & Tobin, 1994), and distortions or displacements of highly regarded curricula and local programs (e.g., Berliner & Biddle, 1995; Meier, 1995; Schaffer, Nesselrodt, & Stringfield, 1997; Smith, 1991).

Nevertheless, in formulating the *Reading First* provisions of NCLB, for example, practicing educators and professional education and educational measurement organizations again played marginal roles (Miskel & Song, 2004). Subsequently and perhaps unsurprisingly, measurement researchers have identified technical and conceptual flaws in NCLB's test-driven accountability plan (Haertel, 2002; Hill & DePascale, 2003; Linn, Baker, & Betebenner, 2002; Shepard, 2002; Sirotnik, 2004). The capacity of flaws to inhibit the achievement of policy aims or to result in unintended negative consequences underscores the importance of determining whether NCLB is reaching its aim of improving education such that no child is, in fact, left behind.

With NCLB slated for Congressional reauthorization in 2007, it is especially important at this juncture for research to document impact in both promising and problematic contexts by investigating questions such as those which motivated the study reported here: *Does local response to NCLB represent successful implementation of the policy? Is the educational impact of NCLB at the classroom level positive?*

This study began with the assumption that NCLB might have both positive and negative effects. In the realization that early implementation can be stressful even for policies that are ultimately smoothly effected and embraced, districts and schools that appeared likely to meet policy requirements, local "best case scenarios," were selected as research sites in an effort, for purposes of policy analysis, to moderate the influence of initial strain. If schools in southwest Washington could successfully achieve NCLB goals, these sites appeared likely to do so.

Method

In situ qualitative study promotes discovery of the details of implementation and impact, revealing policy effects at the classroom and school levels. This type of policy analysis bears a likeness to *utilization-focused evaluation* (Patton, 1997) and *impact evaluation*, as does educational research overall in an era of focus on "what works" (Chatterji, 2005). To determine whether local teaching and learning were improving as intended, classroom observations were systematically undertaken, teachers and school and district administrators were interviewed, and relevant documents and test score statistics were analyzed.

Site Selection and Description

The two focal districts varied along several dimensions (see Table 1). In each district, a school enrolling fourth-graders—one of the three grade levels at which Washington’s standards-based test was administered—was selected and assigned a pseudonym, one school and district referred to as Highland and the other as Riverside. Even in the state which registered the highest SAT scores in the nation (Wood, 2005), both schools matched or exceeded aggregated state scores on the Washington Assessment of Student Learning (WASL), the state’s standards-based test.²

The districts each enrolled fewer than 2000 students, and the two schools’ populations hovered around 500 both years of the study; enrollments grew slightly in year 2. Both districts and both schools enrolled very high percentages of Caucasian students (89–97%), very low percentages of special education students (less than 12%), and relatively low percentages of students eligible for free and reduced lunches (17–38%). One district received Title 1 funding and the other did not. Both districts’ faculties were considered stable, the average years of teaching of the faculty at the two schools ranging from 9–22 years.

The number of participating educators in the study overall was 25. At both schools, additional teacher-participants were recruited in year 2, partly due to attrition (e.g., between years 1 and 2, two-thirds of the participating teachers at one school left for positions elsewhere). The principal at one school retired at the end of year 2, and there was a change in the superintendency in the other district at the end of both years 1 and 2.

² Washington had the highest SAT scores among states in which more than half of the students took the test in 2004 and 2005.

Table 1
Comparisons of Local School Sites with Districts and Washington State

Variable	Riverside		Highland	
	2003-04	2004-05	2003-04	2004-05
School population				
Total enrollment	430	474	468	506
Percentage white	91%	90%	97%	97%
Special education enrollment	7%	12%	3%	12%
Free and reduced lunch	38%	31%	18%	17%
District population				
Enrollment	1,843	1,890	1742	1952
Schools	4	4	4	4
Percentage white	90%	89%	96%	95%
Special education enrollment	10%	9%	0%	10%
Free and reduced lunch	28%	27%	18%	13%
Washington State				
Total enrollment		1,020,959		1,020,959
Percentage white		71%		71%
Special education enrollment		12%		12%
Free and reduced lunch		36%		36%
Stability indicators				
Principal's experience (in years at end of school year)	4	5	3	4
Mean teacher experience, school	10	9	20	22
Participating teachers	3	4	3	2
Participants who left school during or after year	2	2	1	1
Superintendent's experience (in years at end of school year)	2	3	3	1 (interim)
District funding supplement referenda?	None	Fall 2005, failed	None	Fall 2004, passed
State superintendent status	Continued	Relected ^a	Continued	Relected ^a
Percentage of school 4 th graders meeting state standards				
Reading	75.8%	87.5%	89.9%	87.0%
Math	66.1%	73.8%	75.3%	77.0%
Writing	68.3%	57.5%	61.2%	64.6%
Science	30.4%	36.0%	28.5%	43.3%
Percentage of district 4 th graders meeting state standards				
Reading	72.6%	77.9%	79.0%	86.7%
Math	59.0%	62.4%	62.7%	69.5%
Writing	70.8%	60.7%	70.6%	74.0%
Science	40.6%	41.5%	43.5%	49.3%
Percentage of state 4 th graders meeting state standards				
Reading	74.0%	79.5%	74.0%	79.5%
Math	59.9%	60.8%	59.9%	60.8%
Writing	55.8%	57.7%	55.8%	57.7%
Science	N/A	N/A	N/A	N/A

Variable	Riverside		Highland	
	2003-04	2004-05	2003-04	2004-05
Washington State NAEP 4 th graders by achievement level				
Below basic		25%		25%
Basic		41%		41%
Proficient		31%		31%
Advanced		3%		3%
State SAT ranking ^b	1	1	1	1

Source: State assessment website. <http://www.k12.wa.us/assessment>. All sites and participants are identified by pseudonym.

^a Washington Education Association opposed the reelection of the state superintendent in 2004.

^b Ranking of scores among states where more than half of students take the SAT. *Source*: Wood, 2005.

Data Collection

The primary approach to the research was qualitative, subscribing to a view of human phenomena as socially constructed (Vygotsky, 1978) from individuals' perceptions of reality. The research process adhered to interpretive research traditions respectful of emergent design, multiple perspectives, and inductive analysis (Denzin & Lincoln, 1994; Eisner, 1991; Erickson, 1986; Lincoln & Guba, 1985; Wolcott, 1994).

Data collection included 47 observations, 45 of these in classrooms where drop-in privileges were negotiated to minimize observer effects. Semi-structured interviews were conducted (Fontana & Frey, 1994; Rubin & Rubin, 1995; Krueger, 1994), individually with administrators (eight overall) and collectively in focus groups with teachers (four overall). On two occasions, district superintendents elected to be interviewed with either an assistant superintendent or a district assessment officer present. At the end of year 2, a continuing assistant superintendent was interviewed during a period of transition to a new superintendent. Quantitative data collected from documentary sources, primarily in the form of demographics and statistical indicators, augmented observation and interview data.

Data Analysis and Conceptual Frameworks

Data analysis was emergent in character (Denzin & Lincoln, 1994; Erickson, 1986; Maxwell, 1992; Wolcott, 1994). Ongoing analysis using constant-comparative method (Glaser & Strauss, 1967; Strauss & Corbin, 1990, 1994) yielded preliminary interpretations subject to further investigation. Following completion of fieldwork (Wolcott, 1994), thematic content analysis (LeCompte & Preissle, 1993; Miles & Huberman, 1994) was undertaken with reference to three types of analytic frameworks: Bronfenbrenner's (1979) stages of ecological analysis, a sociological model promoting identification of connections between values, organizations, relationships, and practices; Knapp's (1997) articulation of the stages involved in progressing toward educational reform; and three theories of change specifically in response to test-driven educational accountability.

Ecological analysis

The first and most general type of analytic framework used in this study was Bronfenbrenner's (1979) model of human social behavior. *Ecological analysis* promoted

macroanalysis of the ideologies of participants regarding education and educational practice as compared to policy objectives; *exoanalysis* of local organizational and management contexts; *mesoanalysis* of the perceptions and working relationships among teachers and with administrators; and *microanalysis* of classroom activities and behaviors involving or affecting students. Among other things, this model encouraged recognition of inconsistencies between participants' statements of values and observed behaviors or allocations of instructional time.

Stages of reform

The second type of analytic framework used in this study was Knapp's (1997) conceptualization of reform implementation. According to this model, successful reforms pass through four stages: *incremental increase* based on countable indicators suggesting movement toward the reform vision; *professional learning*, including questioning of current classroom practice and getting professional development; *grafting of reform onto familiar practices*; and *full embodiment of reform* (pp. 258-59). This conceptualization offered a continuum for identifying the progress toward reform achieved at the research sites.

Theories of change

The third type of analytic framework responded to Elmore et al.'s (2001) charge that test-driven accountability lacks an explicit theory of change, working from mere presumption that high-stakes testing will improve educational outcomes without posing how or why. Data were analyzed from the perspectives of three theories of change which offered differing explanations of the black-box mechanisms linking accountability policy and desired outcomes.

Standards as effectors. In the first, published shortly before enactment of NCLB, the National Research Council described *standards* as the effector of improved teaching and learning (NRC, 1999) and standards-based tests as documenting achievement gains. Washington's content standards, the Essential Academic Learning Requirements (EALRs), served as the intended basis of the state's test, the WASL, at grades 4, 7, and 10. The EALRs were generally accepted by teachers but not so the WASL (Washington Education Association, 2004).

Similar to the situation in 19 other states, Washington lacked standards-based tests for grades 3, 5, 6, and 8 and, to comply with NCLB's testing requirements for grades 3–8, had filled in the gaps with "off the shelf / norm-referenced test[s]" (Skinner, 2005, p. 86). This study offered opportunity to examine the actual role of standards in classrooms in a standards-based accountability context.

Tests as effectors. The year after enactment of NCLB, Mabry, Poole, Redmond, and Schultz (2003) described a second logic in which *tests*, rather than standards, propel change. In this model, accurate test scores would lead to valid score interpretations (see AERA, APA, & NCME, 1999; Messick, 1989) manifested in appropriate rewards and sanctions which motivate teachers and students toward improved learning. In actuality, frequent scoring disputes or errors have shown that accurate scores cannot be presumed (e.g., Galley, 2003; Hall, McDonald, Scherich, Vickers, & Zebrowski, 2001; Lemire, 1998) and that score interpretation and use may be contested or invalid (e.g., Haney, 2000; Toenjes & Dworkin, 2002). This study included opportunity to examine the function of tests and their consequences in schools and classrooms.

Data as effectors. A third theory of action proposed by Marion and Gong (2003) posited an information loop in which school *scores* reported to parents, the media, and the state generate expectations and pressures and influence resource allocations. Expectations and pressures filter through administrators to teachers, effecting improved teaching and learning subsequently reflected

in improved scores. In practice, aggregation of student scores to the school level presents a measurement issue: few, if any, state tests of student achievement have been validated for purposes other than measuring individual student achievement (Shepard, 2002). This study provided opportunity to reveal whether public reporting of aggregated scores affected pressures on teachers and their classroom practices.

Test score gains were considered in light of measurement literature regarding the tendency of scores to rise gradually and then to fall precipitously with implementation of new tests (Linn, 2000), potentially a result of “score pollution” (Haladyna, Nolen & Haas, 1991) as school personnel administering the tests succumb to pressure to raise scores (e.g., Cannell, 1987; Haney, 2000; Sternberg, 2002). This research also took into consideration the potential for intensified susceptibility to score inflation when policy outcomes are unattainable—NCLB’s goal of universal literacy and numeracy empirically counterindicated by NAEP scores over a 30-year period (Haertel, 2002)—or beyond the control of those held accountable (Harmon, 1995)—NCLB’s Annual Yearly Progress (AYP) requirements described as especially “unrealistic” for schools and districts where small enrollments of identified subgroups of students threaten reliability and statistical stability in aggregated test results (Haertel, 2002; Hill, 2002).

Discussion of findings is organized by five themes that emerged from the data during analysis: local uncertainty during a period of federal policy adjustments and lawsuits challenging NCLB; local compliance and resistance regarding curriculum authority and alignment with standards and standards-based tests; professional impact of NCLB on local educators; disparate conceptions of education in the best interests of children; and validity concerns expressed by participants.

Volatility and Uncertainty

Test-driven accountability in general has been the subject of sustained criticism (Baker, Linn, Herman, & Koretz, 2002; Herman, Baker, & Linn, 2004; Koretz, 2001; McLaughlin, 1991; Rumberger & Palardy, 2005), as have practices derived from NCLB specifically (e.g., Hendrie, 2005b). The volatility of the national policy context during the period of this study is suggested by the requests from 47 states for waivers from their own federally approved AYP plans (Olson, 2005a, 2005c), by state legislation intended to overrule NCLB (Sack, 2005), and by a half-dozen lawsuits filed against NCLB (Archer, 2005a; Hendrie, 2005a; Keller & Sack, 2005; Olson, 2005b).

Within the state of Washington, Superintendent of Education Terry Bergeson, narrowly reelected in 2004 over teacher opposition to state testing, joined a multi-state effort to urge federal policy modification (Bergeson, 2005). Locally, a superintendent in southwest Washington joined a regional consortium which publicly blasted NCLB as lacking “common sense” (Benson, et al., 2004).³

As policy requirements were contested and modified, the administrators participating in this study struggled to understand and meet NCLB’s proficiency requirements and to identify and allocate resources to ensure their compliance. During the period of the study, procedural adjustments to NCLB complicated local forecasting of near- and long-term expectations to be met.

³ The superintendent mentioned was not from one of the districts participating in this study.

Proficiency Requirements: Changes and Cautions

National controversies were often reflected in locally expressed concerns. For example, in year 1 of the study, all participating district and school administrators expressed anxiety about NCLB's "unrealistic" requirements regarding the academic proficiency of students. In year 2, following NCLB adjustments (Davis & Hoff, 2005; Hoff, 2005b; Olson, 2004), some of which had been correctly anticipated by local district administrators, they expressed less concern about reaching NCLB's literacy and numeracy goals:

100% [proficiency]? I don't think that's real.... But will we see more students coming closer to meeting standards? I think so. (Highland district administrator, June 23, 2005)

[Because the state allows some deviations from policy,] we don't have to have all kids performing at standard to *say* we have 100%.... Is it a [foreseeable] reality? I think we can get pretty close. (Riverside district administrator, June 29, 2005)

One principal echoed this qualified optimism: "You're never going to get 100%, but I wouldn't be surprised if we hit 90–95% proficiency" (Riverside principal, June 22, 2005). The other principal expressed less confidence:

We did really well in reading last year, but there were some kids who, despite every opportunity, did not show growth on the WASL.... When you look at all the resources and the wonderful teachers we have, if there's any place it could happen, it would be here. But, for whatever reason, not every child can be at the same place.... I don't know. I don't know. (Highland principal, June 28, 2005)

Teachers also expressed uncertainty, one wondering, "When will the test change to meet the kids? Or will we have to keep bending to meet the needs of the test?" (Riverside teacher, grade 3, May 3, 2005). Teachers' doubts were related in part to their surprising lack of information about NCLB, which a Highland district administrator explained in this way:

[Implementing NCLB] is proving to be fairly difficult. Our issue has been and continues to be a lack of awareness of the law and a sense of urgency... Teachers think it involves other school districts, bigger school districts. (June 23, 2005)

Funding: No Easy Options

Implementing procedures to meet external accountability typically requires new resources, often not easily found or allocated (e.g., Cohen, Raudenbush, & Ball, 2003; Gillborn & Youdell, 2000). At the national level (e.g., Broder, 2005; Gewertz, 2005; Rotherham, 2005) and at the state level, NCLB-related expenditures were raising questions (e.g., Archer, 2005b). In the third year of policy implementation (year 2 of the study), seven lawsuits against NCLB were filed in state and federal courts, two of which challenged the law for violating its own prohibition against unfunded mandates (Anderson, 2005; Archer, 2005a; see also Olson, 2005b).

Reflecting the national context, uncertainty related to identifying or obtaining the funds needed to meet NCLB requirements was cause for local concern. Although both participating districts had recently succeeded in raising funds through elections, one during the timeframe of the study, finding sufficient resources to meet policy challenges remained daunting. A Highland district administrator reported:

My concern is the unfundedness of it. It's unfunded in a very peculiar, particular way. A lot of the [federal grant] funding, such as is available, is driven by

aggregate numbers. In a small district like this, we can't even begin to [apply for] competitive grants. We *have* kids here who need added support, but we have three or six or twelve [eligible kids]. We can't meet the needs test, but the needs are here.... [Using funds from a private grantor,] we put into place things that really helped students and teachers with learning, but that money has gone. We don't know how we can keep on doing that. It's pretty worrisome. (June 23, 2005)

As some states and districts across the country pressed for more policy modifications (Hoff, 2005b), others risked funding cuts through policy defiance (e.g., Hoff, 2005a, 2005c) and "opting out" by refusing Title 1 money in order to avoid NCLB penalties (Davis, 2005; Zehr, 2004). A local district administrator noted the financial difficulties of opting out, even in a district which had successfully passed a funding levy just a few months previously:

I don't know if people are ultimately going to love me or curse me for this, but I've said I think it's time to take Title 1 dollars off the table. We did refuse it for two years, but we need the money, frankly. (Highland district administrator, June 23, 2005)

Insufficient personnel resources were among the specific funding-related issues identified by administrators, including a principal who lamented:

[NCLB] is top-down—"You will do this," "You will do that"—and there's no money to do it. It's like trying to get blood out of a turnip. We're in a small district, so we have small funds and fewer people to do [what NCLB requires].... Our district doesn't have the funds to pay teachers for 2–3 days [of professional development] before the school year starts. (Riverside principal, June 22, 2005)

Fiscal concerns sometimes led to conceptual questions about NCLB's policy framework. "Ideally," an interviewee summarized, "I would only hold people accountable for what could be adequately resourced" (Highland district administrator, June 23, 2005).

Compliance and Resistance

Achievement of the state's own educational reform initiative (Washington State Senate, 1992) had not been fully attained in the decade since its legislation (Borko, Wolf, Simone, & Uchiyama, 2003; Mabry, Poole, Redmond, & Schultz, 2003), suggesting it would not be realistic to expect NCLB's reforms to be fully implemented two or three years after enactment. Evidence of the level of reform achieved locally included "incremental increase based on countable indicators" (test scores) and increased teacher training and "professional learning," corresponding to the three early levels of Knapp's (1997) continuum of classroom implementation of reform: *incremental increase* based on countable indicators suggesting movement toward the reform vision; *professional learning*, including questioning of current classroom practice and getting professional development; *grafting of reform onto familiar practices*; and *full embodiment of reform* (pp. 258-59).

All but one administrator described NCLB as providing leverage to effect needed change without damaging working relationships with faculty by appearing heavy-handed. Although they described NCLB's expectations of universal literacy and numeracy as unrealistic or unreasonable, most administrators indicated that the policy's positive consequences outweighed the negative, one explaining:

I keep calling it the moon shot in public education. You might as well go for a big goal. We're going to be better for having tried. Special education leads the list [of children I'm worried about]. ELL probably comes second. Children of

generational poverty are, in this school district, particularly disadvantaged. The frame of reference here is, “Who are these kids, and what are they doing here?” I worry about those kids, who tend to be invisible. We are now starting to look at what it means to grow up in generational poverty. (Highland district administrator, June 23, 2005)

Students who had been neglected or “invisible” to teachers were being better taught, most administrators said. Their testimony suggested that they were trying to use NCLB to effect reforms they themselves had long considered worthy, reforms to which they considered some teachers resistant.

Curricular Control

Reform not only broadened focus to include previously neglected students but also involved changes to curriculum, pedagogy, and roles. Prior to NCLB, teachers had largely determined the delivered curriculum, often through classroom-level adaptation of district text adoptions or curriculum guides, but external authority was exercising increasing control. Most administrators believed this shift represented educational improvement, and they approved; most teachers did not.

The battle over “fluff”

Most participating administrators reported that NCLB was having a positive impact on instruction through the elimination of “fun projects” of low educational utility. Such projects proved to be contested local ground, most administrators pressing for instruction that could lead directly to measurable gains while teachers defended activities they described as developmentally appropriate and motivational. Administrators reported:

From my standpoint, the overall impact [of NCLB] is positive. I don’t think teachers would necessarily agree with that.... My favorite line has been, “Teach the stuff, forget the fluff.” ... It’s been hard to sit down with a teacher and talk about that when parents like it if the students are happy and having fun in school... [But] if the only thing that’s fun is stuff that doesn’t have any relevance, we’re missing the boat. (Riverside district administrator, June 29, 2005)

Teachers are more accountable for teaching the curriculum, which benefits kids. In years past, teachers taught their favorite things—their dinosaur units—which may have been fun for the kids, but it didn’t really enhance reading and writing.... What we found was that teachers were doing “fluffy” activities that really weren’t focused on learning. (Riverside principal, June 22, 2005)

We are actually moving toward a curriculum that is aligned with our state standards. That’s definitely a positive.... The tradition in this district was that teachers pretty much did their own thing in their own way... Teachers have been getting more precise about setting targets... They’re making those targets obvious to students and then providing instruction that actually helps students to attain the targets. (Highland district administrator, June 23, 2005)

Teachers almost unanimously disagreed, citing “fun projects” as productive for motivation, for positive educational climates and, instrumentally, for learning.

Their ability to determine whether or not to include “fluff” affected teachers’ perceptions of their professional decision-making opportunities. Some teachers expressed fear of being caught engaging students in activities they considered beneficial but not necessarily standards-based, for example:

Today, we were writing plays. We had to [reserve] that [until] after [taking the] WASL.... For 25 minutes, [students] were trying on costumes, showing each other... Now they’re passionate about the plays. But there’s no EALR for feather boas. If someone had come in, what was I supposed to say? (Riverside teacher, grade 3–4, May 19, 2004)

In the struggle to retain such activities, teachers indicated they were losing ground:

You get [students interested] when there’s something that they care about. When we can’t do that, one of our tools has been taken away. (Highland teacher, grade 4, May 2, 2005)

The principals at both participating schools had previously been teachers, but the views of only one corresponded with that of the teachers. The Highland principal explained the utility of some “fluff,” saying that the pressure to eliminate it

... is not really coming from me. However, they do hear comments from the district level. I taught here for ten years before I became a principal, so I’m aware of a lot of the projects. Even I’m like, “Oh, I don’t want to stop doing hot air balloons, you guys!” because I love that [project].... Because of the requirements of No Child Left Behind or the state standards, [teachers] feel they can’t do those things any more without doing a disservice to kids.

Should we throw out some pet projects? ... I have a teacher who goes to Africa every summer. She brings in some very rich cultural things she can integrate with all different areas and do those things she is really passionate about. The kids can learn from that. (June 28, 2005)

At the microsystem level of classroom implementation (Bronfenbrenner, 1979), the inclusion or exclusion of “fluff” affected students’ educational opportunities and experiences. Activities were observed which might indeed have been considered “fluff,” as illustrated by the following observations of two fourth-grade classrooms. Some observations appeared to support teachers’ views about the importance of such activities for building bases for learning, and some exhibited all the ambiguity suggested by the dichotomy of opinion regarding appropriate classroom practice.

Observation, Riverside Elementary School, April 16, 2004. The classroom was noisy and happy-sounding, chairs filled with fourth-graders and kindergartners reading aloud together. From a side table, the fourth-grade teacher offered snacks of raisins, fruit-filled cookies, and pretzels.

Two fourth-grade girls knelt, holding a Dr. Seuss picture book open for the kindergartner seated before them, and took turns reading each page to her. One read fluidly with expression, “We sat in shock and disbelief. ‘Oh, no!’ we moaned, ‘Oh, no!’”

A fourth-grade boy moved his finger across lines in a counting book, pointing to each word as the kindergartner read aloud. The younger boy needed help with most words but gamely persisted, his older tutor readily assisting and correcting him:

“Then,” the kindergartner read.

“No, ten,” the fourth-grader corrected.

“Move.”

“No, more.”

Three girls balanced a tiny picture book open on a desk, touching it only when the book threatened to topple forward, as their kindergartner read the final page confidently, “You can help me. You are not too little to play.” The younger child snapped the book closed, saying, “Goodbye. All done.”

“How did she do?” the kindergarten teacher asked. “Was she a good reader?” The fourth-graders nodded, affirming that she was.

As the reading groups finished, five students were looking at the class crayfish, who obligingly waved its feelers as a boy gently took it out of its tank.

Two girls allowed the class turtle to walk around the carpet.

The fourth-grade teacher of this classroom expressed apprehension that her students, acting as reading tutors, were not engaged in what external authorities might consider appropriate reading instruction at fourth-grade level. She herself, however, felt such activities were advantageous for her students, especially for promoting the competence and confidence of struggling fourth-grade readers, and supported by research (see Green, Alderman, & Liechty, 2004; Kerr & Verhaeghe, 2005). She explained:

If you have a child who feels socially and academically incompetent,... [She or he is] not going to be as successful as a student who feels, “I’m going to share this, and I’m not going to be laughed at.”... There’s not a focus on that at all, except in what the *research* is saying but the *state* isn’t. At least, I don’t see it. (May 19, 2004)

Another teacher presided over an activity with less clear connection to learning or reform initiatives:

Observation, Highland Elementary School, grade 4, March 23, 2004. Two boys carried car parts featuring specialty painting into the classroom. Jeri, the guest speaker, followed them, greeted the children, and explained that she had once attended their school and that their teacher had once been her teacher. “How many of you like cars?” she asked. Hands shot up. “Anybody not like cars?” Two hands were raised amid general laughter.

Jeri exhibited posters, a paint gun, painted car parts, and an “airbrushed” mailbox with flames spreading backward from the mail opening. She explained how she had sanded and painted each piece. There were “Ooooooh”s of interest when she mentioned she also painted Harley-Davidson motorcycles. She described the training she had undertaken, confiding, “I was the only girl in 58 people.”

“Wow!” the teacher said. “Aren’t we proud!” The class broke into spontaneous applause. The teacher encouraged the children to think about their future careers while she turned on a promotional video Jeri had brought.

After showing the video, both women encouraged the girls in the class to consider traditionally male-dominated careers, Jeri describing herself as “living proof” that women could succeed and enjoy such occupations.

No clearly academic instruction took place during this observation. Had she been challenged, the teacher might have justified the time allocated to this activity as career education or a proactive countering of gender stereotypes. That she was feeling less free to attend to such things was indicated in a later interview:

There’s a rigidity I’m feeling now that I didn’t feel about four years ago. It’s coming from the demands from the state as to what’s required for us to accomplish with our students. When I came here, we had our own curriculum

and, as a team, we determined what was best for students. Now, guidelines are set. (Highland teacher, grade 4, May 2, 2005)

Curricular alignment and other “test prep”

For coherence, standards-based testing systems rely on alignment of both curricula and the tests to content standards. Aligning curricula with content standards was described by local educators as an appropriate route to improving scores. Few teachers opposed the EALRs, and several praised them, suggesting substantial local compliance with the content standards. One teacher said, for example:

It’s been valuable, the things we’ve been going through the last ten years. We’ve improved the curriculum, which was very valuable. We’ve really improved our writing instruction. The EALRs and all this stuff has been very helpful. (Highland teacher, gr. 3, May 2, 2005)

A colleague offered a quick caveat: “The guidelines are quite good, but the pressure that comes on teachers and on students is not” (Highland teacher, gr. 4, May 2, 2005). A few teachers indicated that the content emphasized by the EALRs was appropriate, one describing the authenticity of the math curriculum:

I’m teaching connected math, which is applicable to problems the students will have in real life. There are no drill-and-kill sheets saying, “Here’s the algorithm. Now, do it.” Instead, they explore, like, “Here’s a strategy. Aha!” (Riverside teacher, gr. 6, May 3, 2005)

Most teachers, however, expressed concern about the overwhelming numbers of content standards, the Grade Level Expectations (GLEs) that elaborated the EALRs, one reporting: [With] the sheer number of [GLEs] I’m expected to teach... if you don’t start in the first week and just go for it, you simply cannot teach to all of them....

Teaching this past year was like... a zoo, a boring lecture with no art until after the WASL, working endlessly but getting nothing done.... I feel like I taught the [GLEs] and worked on classroom management all year. (Riverside teacher, gr. 4, May 19, 2004)

This teacher’s reference to a lack of art instruction reflected a general worry among teachers that untested subjects were being squeezed out of the curriculum. The fine arts, social studies, and sometimes science were described as threatened, for example:

I came from a school in Massachusetts where we had a strong integrated social studies program that drove the whole year. Here, when the topic of social studies came up, it was [suggested that] I could do it after the WASL. Whoa! That just blew my mind! (Riverside teacher, gr. 4, May 19, 2004)

Teaching to the test itself, rather than to the standards, was uniformly perceived to be inappropriate, although no interviewee suggested that aligning the curriculum to state content standards might have the effect of pervasive teaching to the (presumably) standards-aligned test. Not all parts of the WASL (i.e., not all subjects at all grades) had been subjected to external examination to determine alignment to the EALRs, leaving open the possibility that a teacher’s implementation of a curriculum aligned to the standards might not result in her students’ preparedness for the WASL’s content.

Given such unresolved tensions, classroom observation data unsurprisingly indicated neither total policy compliance (e.g., uninterrupted attention to tested topics) nor total resistance. Some “test prep” was clearly observed, as indicated in the following vignette.

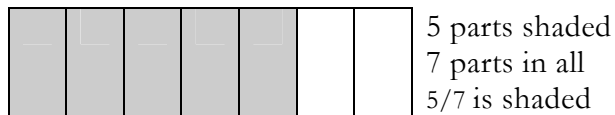
Observation, Riverside Elementary School, grade 4, April 16, 2004. As the observation began, the teacher hurriedly explained, “This is not our usual seating arrangement,” referring to a grid of student desks all separated from each other by rows and aisles. “It’s just to get the kids ready for WASL testing next week.”

The daily agenda posted at the front of the room indicated that the first four hours were to be devoted to academic areas related to the WASL and NCLB—1.75 hours to math (math and math journal), 1 hour to reading (independent reading and reading aloud), and 1.25 hours to language arts (cursive writing, a spelling test, and a writing assignment).

Although no other academic areas (e.g., science, art) were listed, several maps on display indicated that the teacher did give attention to social studies. Half of the classroom was encircled by an historical timeline near which 3x5” index cards indicated milestones. One card, for example, offered information in a child’s handwriting: “1732 George Washington is born” with a drawing of a bottle-wielding baby in a crib. Another read, “1835 Abner Doubleday invents baseball” and featured a drawing of a man and a floating ball.

In this setting, students were working on math worksheets with WASL-like fractions problems such as:

Parts of a Whole



number of shaded parts → 5 ← numerator

number of parts in all → 7 ← denominator

The teacher fielded questions from the students, circulating among them to help individuals. Occasionally, a student asked a question aloud, and the teacher gave a general explanation, sometimes writing information at the board.

After huddling at a desk near the front of the room, two girls excitedly brought their work on a math problem to the teacher’s desk. One danced in place, pumping her fists in the air. When the teacher confirmed their solution, they crowed, “We did it! We did it!”

In other observations, odd-fitting connections between classroom activities and the state test suggested Knapp’s (1997) “grafting” of new onto old practices, as indicated in the following vignette.

Observation, Highland Elementary School, grade 4, April 16, 2004. The teacher appeared to be working with students on several things simultaneously. As flowers were projected on a screen at the front of the classroom, the teacher quizzed students, asking them to identify the flowers in writing.

Part two of the quiz followed, the teacher distributing to each student a paragraph about wildflowers. The teacher directed students to insert needed punctuation into the paragraph, then noted to the observer, “This is one thing we do to work on writing mechanics.”

Although observations such as these revealed a mixed picture in terms of classroom devoted to test preparation, interviews suggested that teachers were acutely conscious of expectations related to the state’s standards-based test. One teacher worried aloud:

I feel I have to be accountable for [having] every minute of the day match some kind of EALR.... The rush to get through the EALRs—there’s so much pressure!

It's like a checklist of all these things without any depth.... Here, we talk about WASL all year long.... It's not "How are we making our kids passionate about learning?" It's about whether they're passing the WASL. It's "What is your percent [of students scoring at the] proficient [level]?" (Riverside teacher, gr. 4, May 19, 2004)

Washington's hybrid testing system administered the Iowa Test of Basic Skills (ITBS) and Iowa Test of Educational Development (ITED) to students in non-WASL grades. Teachers indicated that pressure was most intense at grade levels targeted for the WASL:

A lot of the pressure is on the fourth-grade level. [In third grade,] I have a lot of freedom to teach in ways that I think are best for the kids.... For example, we're using [a commercial] reading curriculum which doesn't go with the best practices in literacy that I've learned. I feel I have the freedom to figure out ways I could use that curriculum. I won't as much next year because the third grade will be required to take the WASL. (Riverside teacher, gr. 3, May 3, 2005)

As suggested by the NRC theory of change (NRC, 1999), state content *standards* were indeed influencing change in what was taught and learned in classrooms in the two schools participating in this study. In large part, however, the standards appeared to be influential because the state's standards-based *test* functioned as an enforcement lever, as suggested by the Mabry, Poole, Redmond, and Schultz (2003) theory of change. That is, evidence from this research suggested that it was an intersection of these two models that drove classroom change. Whether the change constituted reform—in the sense of radically improved teaching and learning—remained a clouded issue with local proponents passionately arguing each side.

Professional Impact

Data suggested that the reshaping of professional roles and responsibilities, begun with the enactment of Washington's standards-based education reform act (Washington State Senate, 1992), had intensified under NCLB.

Teacher Autonomy

Federal determination of subjects to be prioritized (i.e., reading and math) and state determination of content to be taught (i.e., the EALRs) were justified locally by a district administrator:

We are actually moving toward a curriculum that is aligned with our state standards. That's definitely a positive. By doing so, I think we have taken steps toward making the educational experience in this district more equitable. The tradition in this district was that teachers pretty much did their own thing in their own way, isolated from each other, and it was possible for kids to have an extraordinarily fine and rich experience with a particular teacher [with an] inventive curriculum, but pity the kids whose parents did not come to the schoolhouse door and say, "I want *that* teacher." (Highland district administrator, June 23, 2005)

Other administrators explicitly approved NCLB's moves toward teacher accountability, one arguing that neither teacher professionalism nor autonomy were threatened:

I've been working in state schools since the 1970s. Throughout that time, teachers have looked at the profession as *their* profession, and they don't want somebody else controlling it.... I'd argue NCLB doesn't have any effect on your teaching. No one is walking around from the federal office to check what you're teaching.... I think they have a lot of academic freedom to teach, but it has to be put into perspective. They're employees, not individual contractors. We pay teachers a good wage, and they should be accountable, not doing something just because it's fun.

They're professionals. Being professional is also about taking pride in a job done well. I can cite thousands of examples where they're not demonstrating the professionalism they say they deserve.... Teachers here will tell you they don't have enough input. I'm not buying that story because they have opportunities to get involved, [like] participat[ing] on our School Improvement Plan, but they're not stepping forward or they're not doing it unless they get paid for it. (Riverside principal, June 22, 2005)

Another approved the accountability system's power to strengthen teachers' expertise:

This district is all about achievement, what happens with learning in the classroom. Teachers have to be accountable—"If a child isn't getting the concept, what are you going to do about it?" It makes them be more precise. Their repertoires may have to grow; they may need to collaborate as a team; the team may have to act as a microcosm; they may have to do more communicating—and they are. (Riverside district administrator, June 29, 2005)

One of the two principals, expressing confidence in teachers, had reorganized the school week to promote collaborative decision-making by grade-level teams. This principal struggled with a desire to buffer teachers from stress as opposed to passing pressures and expectations on to teachers, as described in the theory of change by Marion and Gong (2003):

The teachers in this building are, I think, a little unique.... Of 22 teachers, I think three have less than six years' experience.... Our teachers [have always been] dedicated to the kids.... We have staff meetings every week. [This year, each grade-level team also has a] weekly collaboration. I wanted each team to [have more time to] collaborate and be very intentionally strategic about how they use their time. I know how much stress they put upon themselves to make sure they're reaching all the standards... I am torn between feeling I want to take stress off the teachers and do what I can to support them so they can do their best for teaching the kids [but]... sometimes I feel like I should be doing more, pressuring them more. (Highland principal, June 28, 2005)

To teachers, the pressure was already palpable. Pressure at fourth-grade level, where the WASL was administered, was great enough that Highland's fourth-grade teachers had petitioned their district for relief:

The fourth-grade teachers met with the superintendent and principals. Most of us cried and said, "You can't do this to us any more."... Not until we met with the administrators did we feel we were getting support... I think they finally understood what we were going through. They were looking at the test, not looking at the people. (Highland teacher, gr. 5, May 2, 2005)

In addition to individual stresses, some teachers connected pressure with the good of the school community, for example:

I know that there's funds attached to [WASL scores which] are meaningful to the whole school. So I feel like I have to teach to it. I guess there is a lot of pressure.

(Riverside teacher, gr. 4, May 3, 2005)

Some administrators anticipated that NCLB would nudge from the profession teachers who resisted local policy implementation efforts, and data did suggest that stress was triggering flight responses. Each case of participant attrition among teachers between years 1 and 2 of the study involved fourth-grade teachers: two left their schools, two changed grade levels, and one explained, "I moved from grade 4 to grade 5 to escape the WASL, but [with NCLB,] I'm back in it again" (Highland teacher, gr. 5, May 2, 2005). Another fifth fourth-grade teacher retired earlier than planned at the end of year 2. In addition, a third-grade teacher had considered but rejected an opportunity to move to fourth grade, explaining, "I don't want to do it because of the pressure" (Riverside teacher, gr. 3, May 19, 2004).

Despite some personnel losses and other concerns, most participating administrators described NCLB's impact as basically positive. From a policy perspective, the shifts they described suggested progress on Knapp's (1997) scale of effective reform, change from incremental change and conglomerates of old and new practices toward full reform.

On the other side of the personnel line, most teachers described the shift as infringing on classroom decision-making and adaptation, with disempowering and discouraging professional impact. Said one:

[I'm] a younger teacher with 25 years to go before retirement. The pressure of the whole profession, the parents hammering at you, the kids so needy—I can't do this for 25 years.... It's hard. I see myself burnt out in 10 years, bailing completely.... That first year, it was nice to have those [state] guidelines, so I knew what to do.... Now, I've gotten myself into a pattern. There are things I would like to do, but I can't. (Highland teacher, gr. 4, May 2, 2005)

Most teachers resisted not the content standards themselves but the relinquishment of authority to the standards that implementation of standards-based reform, as they experienced it, was effecting.

Data-Driven Administrators

NCLB had altered the responsibilities of district and school administrators. Most notably, local data collection, use, and reporting had increased: "While some of the data is not necessarily applicable, it has helped to create a new kind of environment in which we are actively seeking data" (Highland district administrator, June 23, 2005). All but one administrator expressed satisfaction with school improvements resulting from stronger focus on data:

The data collection [for NCLB] has been helpful in creating change. We have used the data to implement our goals each year.... That data is dissected by our staff, me, and the school improvement planning team, to see how we are making a difference. We try to link teacher goals to the school goals. Every trimester, each teacher sets goals in reading and math according to where we see gaps in the data.... The accountability piece comes tomorrow when we report our goals and their measurement to the [School] Board. (Riverside principal, June 22, 2005)

In addition to coordinating school and teacher goals, district administrators reported that data promoted improved curriculum and personnel decisions:

The data helped me make the case on behalf of children that, in this district, we need to adequately resource reading and writing instruction. For example, last

year, I was able to advocate successfully for the adoption of a core reading series plus a variety of leveled readers for take-home—a great many materials. I also advocated for a literacy coach, actually two, one of whom has worked out very well.... Those have been positive changes that have occurred fairly directly from the new federal mandate. (Highland district administrator, June 23, 2005)

I never thought I'd be the type of person who'd say I love data, but it's helping us make decisions that are better for kids... It's providing supporting information to plan programs for students. Some of our decisions have been reaffirmed by the data [but not all.]... We'd been using SSR—Sustained Silent Reading—for a long time [with] the idea that, if you get kids reading, they'll be better readers, [but] the data didn't support that. Now we're more conscious about whether we're giving them skills, helping them practice appropriate skills. We are looking at the data—scientifically based data—as we make selections for the future. (Riverside district administrator, June 29, 2005)

The teacher-oriented Highland principal appeared as an outlier in describing classroom-level data as more useful than test scores, explaining:

[If I were to choose between WASL data and teacher-generated data to see how students are doing,] I'd go to the teachers.... We're teaching a child. You take out that human piece at times when you just look at data. (Highland principal, June 28, 2005)

One local administrator described NCLB-reported data as useful for clarifications to parents:

It's especially helpful in communicating to parents, in helping them look past their emotional feelings about their children, to share data showing measurements.... NCLB has helped us communicate what *your* child is learning. (Riverside district administrator, June 29, 2005)

Another administrator worried about data that was overwhelming or that eluded clear interpretations:

I try to look at where we are with the WASL. I try to look at “Why are we here? Why did we go down from last year?” We had 89.9% meet standards, so we did fairly well in reading last year. We did try to look at why those 17 students who didn't meet standards didn't meet standards. When we pulled [their records], several of them were on an IEP or they were ELL. We tried to look at what kind of services we were giving to those kids.... It's a struggle. I constantly think I should be reading more data. (Highland principal, June 28, 2005)

All participating administrators, even those who extolled the usefulness of NCLB-required data, reported that the benefits came at a cost. One described administrator flight in advance of 2008, when Washington's graduation test was slated to determine eligibility for high school graduation, and 2014, when NCLB's 100% proficiency in reading and math requirements at grades 3–8 were to be achieved:

If you get a group of senior administrators together, often the first joke is, “So, are you *pre*-2008 or *post*-2008?” or “Are you *pre*-2014 or *post*-2014?” It's taking a big toll... I believe in the standards, absolutely.... I don't want to go back, but it is wearying. (Highland district administrator, June 23, 2005)

The Highland superintendency did change hands in both years of this study.

Public Pressure

“When it comes out in the paper, holy cow!” a teacher exclaimed (Riverside teacher, kindergarten, May 3, 2005) in describing the personal impact of public reporting. In interviews, teachers identified media attention as sources of personal stress less frequently than conversations with parents, although the parents might partly have been responding to media alerts. An administrator confirmed a pervasive general effect:

Teachers can’t get away from knowing NCLB is out there because, every time you turn around, you hear it. At some point, it *is* to excess... As a staff, I think the teachers are worried. It’s always in the backs of their minds, especially the fourth-grade group. (Highland principal, June 28, 2005)

Overall, observation and interview data suggested that scores, reported for government accountability purposes and disseminated by the state and districts to the schools, were stronger influences on local practice than media reports. In this regard, this research elaborates the theory of change articulated by Marion and Gong (2003) by suggesting that different components in accountability systems’ information loops are, in reality, differentially weighted in terms of their effects on outcomes. In addition, the trickling down of information from government sources to public outlets (e.g., news media, school report cards) suggested government’s opportunity to shape information for public consumption, a mediating influence not clearly registered in that theory.

In the Interests of Children

Analyzed with the assumption that all parties shared a desire to provide “what’s best for kids” (e.g., Beach, et al., 2003; Behuniak, DeVito, Rivera, & Fremer, 2001), the data revealed bright-line ideological distinctions. As teachers and administrators described and rationalized different approaches to curriculum and pedagogy, their justifications exposed ideological conflict at the macrosystem level (Bronfenbrenner, 1997). Both teachers and administrators invoked science (i.e., data or research) to support their views.

While NCLB suggested policy-makers’ appreciation of generalizable *systemic change*, teachers’ descriptions of their responsibilities suggested appreciation of improvements resulting from *individual adaptation*. Against NCLB’s pursuit of large-scale “research-based” and “scientifically based” curricula and practices, teachers, in effect, argued for locally developed programming informed by knowledge of contextual and individual student circumstances.

Teachers expressed worry that developmentally appropriate practice, allowing for tailored educational delivery to students, was increasingly threatened by systemic reform initiatives. The magnitude of the inappropriateness, in terms of difficulty level that they perceived, was suggested by a teacher who complained:

We have parents who couldn’t answer the questions on the fourth-grade test....

There’s a developmental inappropriateness with WASL-type questions. (Highland teacher, grade 4, May 2, 2005)

Most but not all teachers who remarked on the developmental appropriateness of the WASL expressed negative opinions, especially those preparing students to take the WASL, as indicated in this exchange between colleagues:

Students are expected to write a five-paragraph essay for the fourth-grade WASL.

So, gradually, we work as a team, coming together to try to find a way to address

these developmentally inappropriate expectations. We have to make it work.
(Riverside teacher, gr. 4, May 3, 2005)

I disagree. They can rise to the five-paragraph essay. (Riverside teacher, gr. 6,
May 3, 2005)

Teachers often articulated complex understandings of student success that were not necessarily measurable and that tolerated divergent outcomes, for example:

I think of success [being defined] in two ways—meeting the EALRs and being where you need to be grade-level-wise.... But [a student] could [alternatively] be academically successful in a developmentally appropriate way. I have a student who is below grade level in skills but has grown this year, and that's academic success, too. (Riverside teacher, grade 4, May 19, 2004)

At these sites, teaching and learning were occurring within an ideological conflict involving scale, the accountability system focused on large-scale improvement and educators focused on improvements within their spheres of professional influence. District administrators focused on district improvements, principals focused on school improvements, and teachers focused on individual student growth. Interviewees' perspectives about whether educational accountability was serving the interests of children was largely a function of their positions within the system.

Each scale level—that of the policy-maker, district administrator, school principal, and teacher—engendered its own particular interest concentrations, each suggesting different values. At the same time, each scale level was interconnected with the others, such that pursuit of the values associated with one level could be inhibited or prevented by pursuit of any of the others, suggesting the power of each ideology to threaten the system. Externally mandated and enforced through locally significant rewards and sanctions, the values implicit in NCLB were gaining precedence over those emanating from participating classrooms. If indeed NCLB was “more than an act—it's an attitude,” as declared by U.S. Secretary of Education Margaret Spellings (National Public Radio, January 31, 2005), it was an attitude more consistent with that of administrators than with that of participating teachers in this study.

Student Anxiety

A month before the WASL would be administered to fourth-graders, a third-grader surprised her teacher by describing her test anxiety in a show-and-tell session described in the following vignette.

Observation, Highland Elementary School, grade 3, March 21, 2004. After lunch, children seated themselves in a circle on the classroom rug and shared moments from their out-of-school lives: a family building project, an annual camping trip, a ferry ride, a playground altercation. One boy said, “I'm not going to be here in April, but I'm going to be really sad because I'm still going to have to take the WASL.” Although the WASL would not be administered to these third-graders until the following year, his comment sparked an excited response from a classmate.

“It's a really, big, big, big test!” a diminutive blonde girl exclaimed. Her excitement was contagious, and a small hubbub of chatter erupted.

“One at a time,” the teacher said.

“It's like the biggest test of the year!” the girl went on emphatically.

“Big meaning *important* or big meaning *long*?” the teacher prompted.

“It’s important *and* it’s long!” the girl reported. “It takes like two weeks or three weeks to finish it—I don’t remember how long, but not one week. It takes more than one week. The WASL test,” she paused, dropping her hands in her lap. “I mean, *everybody* in my class at my other school was scared of the WASL test! And everybody didn’t want to move up to the next grade where they’d have to take the WASL test!”

“My sister had to take it three or four times already,” a boy said somberly.

“And did she survive?” the teacher asked hopefully. The boy nodded.

“My brother had to take it a bunch of times,” another boy said. “His brain is rotted.”

Many teachers reported that their students experienced test anxiety. In a group interview, two teachers broke down in tears as they reported:

Boy, do they ever feel that pressure on the WASL! They ... sit down before those blank pieces of paper.... I always have a handful of very anxious students. Last year and this year, a student was getting counseling and losing sleep.... In college, they teach you to scaffold learning... But you push them into the WASL, and [she weeps]... you’re not even allowed to say, “You’ll be OK.” That’s what teachers do ... But, with the WASL, we’re asked to sit and watch them struggle. (Highland teacher, gr. 4, May 2, 2005)

We know, in the long run, they’ll do fine, [she weeps] but we’re torturing them along the way when we could be encouraging them. (Highland teacher, gr. 3, May 2, 2005)

Several described their efforts to minimize student anxiety, efforts they considered largely futile given the constraints of external directives. One said:

This year, [my students] were pretty stressed out. I tried to defuse it. I got them little incentives, little treats. We had a conversation about “What does smart mean? Does it mean you are better person if you can pass a test?” [But] I had kids in tears both last year and this year.... [If they don’t understand one part of the test,] you can’t really explain it to them. They’re just sitting there, like, “Help!” and there’s nothing you can do about it. (Riverside teacher, gr. 4, May 19, 2004)

A fourth-grade teacher described how she had down-played the importance of the WASL to protect her son from negative psycho-emotional effects:

As a parent, I chose not to let my child know his WASL score. He failed the writing part of the WASL. Of course, he asked, “How’d I do?” What should I tell him? He’s a straight A student at the middle school, and it’s not an easy curriculum there. I didn’t want to tell him he was a failure or that he didn’t pass it. I felt he would have given up on writing. He writes to the point. He doesn’t elaborate, and that doesn’t give them what [the scorers] want. (Highland teacher, gr. 4, May 2, 2005)

Although few administrators mentioned test pressure on students, teachers’ reports were confirmed by a principal, who elaborated with information about how school resources had been redeployed specifically to combat test-related pressure:

We had a couple of students this year really kind of break down. They were worried. One little boy thought his teacher was going to lose his job if he didn’t do well on the WASL. There were a couple of fourth-grade classrooms where we

heard that. I don't know if the parents had read something, but it didn't come from the teachers. These students were worried and upset....

The pressure [on the teacher] transfers to the student, which concerns me.... Let me give you an example [of how we tried to handle that].... [During test administration,] we pulled teachers [from non-WASL classrooms] and [had them] as extra support persons in the [WASL] classrooms because the [fourth-grade] teachers are so frazzled.... I'd rather get a sub for a third-grade teacher or a fifth-grade teacher at this point. (Highland principal, June 28, 2005)

This voluntary added expense represented a conscious decision to temporarily risk the educational quality in non-fourth-grade classrooms to combat WASL anxiety in fourth-grade classrooms.

Some parents were reportedly concerned about the effects of test stress on their children, said teachers who shared their concern. One noted:

One of the things I have heard a lot in the community is parents being really upset about the stress their kids are under—second grade, first grade. I had a conversation with a woman at the coffee shop, and she said her second-grader knows the term [WASL and] is scared about it.... I had one kid who was really scared about taking the [state-required] ITBS. We talked about what are tests for,... made the ITBS a game.... That wouldn't work in fourth grade. They know the WASL is the big test. (Riverside teacher, gr. 3, May 19, 2004)

Teachers indicated that some parents had become alarmingly over-focused on test scores, increasing the pressure on their children. One reported:

I have a [third-grade] girl who cries if she doesn't get 100% on all her papers. Her parents make her practice so she can do well on the test.... At parent conference time, we had another mom who was not interested in talking about how her kid was doing. Instead, she wanted to know, "What's going to be on that ITBS?" and how was I going to get him ready. (Highland teacher, gr. 3, May 2, 2005)

Teachers described their efforts to reduce parent anxiety as no more successful than their efforts to reduce student anxiety, saying, for example:

We still can't get parents [of our elementary students] to believe that, if [students] fail the WASL, they will not be retained [at grade level]. We tell them every year, but the parents think they will fail at every grade level if they don't pass the WASL and maybe the ITBS too. (Highland teacher, gr. 4, May 2, 2005)

I just about died one day when a kid in my class, who was taking the ITED, asked, "Is this going to go in my permanent record?" Apparently, he'd been talking to his mom, and they were scared. (Highland teacher, gr. 3, May 2, 2005)

It was the personal effects on children of the WASL, not yet a high-stakes test for fourth-graders, that worried teachers most, some to tears—the impact on self-esteem, motivation, and aspirations. One teacher pointed out:

Think of the child. You're talking about him personally: "Here's your score and you failed." Maybe it was an unrealistic expectation set for him. Maybe it was not even fair to ask that child to do what we asked him to do, but now he knows he's a failure.... I've given lots of standardized tests but, for fourth grade, there is something about [the WASL] that is more difficult. (Highland teacher, gr. 4, May 2, 2005)

Teachers' Goals for their Students

Most teachers considered the stress they perceived in their students to be discontinuous with their own teaching goals and priorities. More than their goals regarding student achievement, teachers' goals regarding student self-esteem, motivation, and aspirations were at odds with the test-focused accountability system implemented by the state and reinforced by NCLB. They noted the conflict explicitly, one saying:

“If you have a child who is passionate about learning, you can build on that. I feel the EALRs put a little bit of a block on being able to do that” (Riverside teacher, gr. 3, May 19, 2004).

Some teachers observed that some of their goals and their work with children was beyond the measurement capacity of the state test, one teacher commenting:

I'm excited my students actually learned. I could see growth from beginning to end. I think [it's because of] all the work we did as a community. We worked really, really well. We're a very close group.... It was my main objective that we would treat each other with respect, help each other. They came in as a very needy class—lots of home issues, not a lot of support at home—so I wanted to make sure that they supported each other....

They came in very low but, when we just did the STAR reading test to see where they were, I had only four students below fourth-grade. When they came in, I think I had ten. They're getting division, and they know all about Lewis and Clark—not that that's on the WASL—and they've matured as individuals, problem-solving their little problems. (Highland teacher, gr. 4, May 2, 2005)

Others noted two chilling effects of testing's limitations: narrowing of the curriculum to tested subjects and lowering of children's self-esteem and motivation when their favorite subjects were lost. Teachers said, for example:

My main concern is the emphasis our school puts on academic success in math and reading. Everything else is secondary to that. That says academic success is more important in those two areas than the others, whereas my opinion would be that they're all at the same [level of importance]. (Riverside teacher, gr. 4, May 19, 2004)

Three of my kids had sort of checked out already in fourth grade. Two of them are fabulous artists, but there's no art [specialist] in this school. And, with the worries of WASL and trying to get things done, I didn't do much art. But if I had, it would have been an area in which they could have been more successful, and other kids could have seen them as successful. (Riverside teacher, gr. 4, May 19, 2004)

Validity Concerns

According to the Mabry, Poole, Redmond, and Schultz (2003) theory of change, test scores would lead to valid score interpretations which result in appropriate rewards and sanctions and improved teaching and learning. However, local educators expressed concerns in non-technical language about the validity of actual uses and consequences of test scores (see AERA, APA, & NCME, 1999).

Teachers expressed concerned that testing was damaging the educational system by squeezing out some subjects and pedagogies, a matter of *systemic validity* (Frederiksen & Collins, 1989). Formative feedback for improving the system did not fully compensate for this damage, partly because of timing: “It’s frustrating partly because we don’t get WASL results until Fall when we’re into another school year” (Highland principal, June 28, 2005).

Some teachers expressed *ecological* or *instructional validity* concerns regarding discrepancies between the ethos of instruction and the ethos of testing. “At this age, writing really requires scaffolding”, said one (Highland teacher, gr. 4, May 2, 2005), noting that teachers were forbidden to provide scaffolding during test time. A colleague illustrated the discontinuity with mastery approaches and encouragement:

In the classroom, they get more chances. You can say to a student, “I’ll work with you. You’ll get it. Oh, look, you’ve got two of them right today!” On the WASL, you don’t have the chance to say that. (Highland teacher, gr. 4, May 2, 2005)

There were local concerns regarding *construct validity*, whether the WASL was measuring what it was intended to measure. Some teachers explicitly doubted whether the state tests really measured their *intended constructs*, for example:

The [new WASL] science test has nothing to do with science instruction. It’s an intelligence test in disguise.... Their level of psychological maturity, their attention problems... there are lots of other things being measured. (Highland teacher, gr. 4, May 2, 2005)

In addition to measuring intelligence rather than achievement, teachers suggested that the WASL actually measured another *rival construct*—student motivation to perform well on the test:

There are students who know the content but just don’t care, and they are very capable of giving you nothing. (Highland teacher, gr. 4, May 2, 2005)

I have really, really good writers, but the prompt didn’t motivate some of them.... I have a girl who said she was sick and tired of doing this test so, on the math test, she wrote, “I don’t know how to do this” for every question. She was just worn out.... I observed one boy, who had just finished the sixth grade STAR test at the 6.0 grade level, looking at a WASL question that demanded inferences and then turning in an almost empty test booklet. I know he’s not performing below grade level, but his score will show that he is. (Highland teacher, gr. 4, May 2, 2005)

Teachers also expressed concern about test scores that reported status but not progress, one saying:

I don’t know what my WASL scores are going to show. In writing, I’d be surprised if any of my students passed. But, from what they were doing at the beginning of the year to what they can do now, that’s OK with me. (Highland teacher, gr. 4, May 2, 2005)

A principal who, when asked about valid data sources, said, “I’d go to the teachers” (June 28, 2005) rather than to test scores in order to understand a child’s achievement, suggested that classroom assessments—formal and informal—provided better information than the state test. A teacher elaborated this point, simultaneously suggesting that the state test provoked suspicion: “You can tell a lot with classroom assessments, and there’s no trickery. But not everything they’ve learned shows up on the WASL” (Highland teacher, gr. 4, May 2, 2005).

Issues of *consequential validity*, “the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores” figured prominently in interview data (Messick, 1989, p. 13; emphasis in original). Data from teachers strongly indicated that most felt the consequences of testing borne by children were excessive and inappropriate.

The appropriateness of the consequences to schools was also challenged. The validity of inferences of school quality based on WASL scores aggregated to the building level, as required by NCLB, was as much in doubt among teachers as among many researchers (e.g., Shepard, 2002). One participating teacher described collective teacher resistance to a district interpretation of her school’s WASL scores:

[We were told,] “You went down in all three [state-tested] areas.”... We had actually made a tremendous jump—we were up 14 points in some areas, down just one point in one area. We blew a gasket. They didn’t listen or even acknowledge the improvement we had made the year before because we were down that one point. (Highland teacher, gr. 4, May 2, 2005)

Teachers’ intuitions about the invalidity of judgments based on test results fed resistance to the test and desire to protect their students and schools from what they perceived to be unfair consequences.

Conclusion

Test data indicating that the two participating schools were meeting state and federal achievement targets suggested successful policy implementation. However, the fuller picture of implementation provided by classroom observations and interviews of personnel was much less clear about whether local implementation of NCLB had been successful.

Test scores did show “incremental increase based on countable indicators,” and observations did reveal that both the state content standards and the state test were influencing curriculum and pedagogy. However, observations indicated more “grafting” of old onto new classroom practices than “full embodiment of reform” (Knapp, 1997), perhaps indicating a predictable movement through phases of policy acceptance or perhaps, alternatively, indicating that most teachers were defying policy where they could and complying where they could not.

Whether the “grafting” constituted improved teaching and learning was a matter of local contention, administrators often considering the movement toward a more standards-based curriculum beneficial and teachers almost unanimously worrying about lost subjects and pedagogies and the erosion of their opportunities to devise or adapt teaching to their particular students. The grafting was observed to produce uneven and odd-fitting conglomerates of practices suggesting unresolved ideological conflicts (see Bronfenbrenner, 1979). While some curricular effects were praised, attention to untested subjects (e.g., art, social studies) was diminishing. Teachers reported that standards and testing were overwhelming what they considered to be best or developmentally appropriate practices.

Interview data revealed that, for participants in both administrative and instructional positions, NCLB’s impact was problematic. At the district level, worries about unrealistic student achievement targets were giving way to acceptance of policy adjustments that would probably allow continued demonstration of acceptable progress. Administrators expressed satisfaction that historically low-performing students, previously ignored, were beginning to receive appropriate educational services. At the same time, NCLB requirements and sanctions were sufficiently harsh that, despite constrained resources, one local administrator had recommended refusing Title 1

funding to escape them. Small districts' ineligibility for federal grant assistance exacerbated local financial burdens.

At the school level, teachers registered more anxiety about the impact of test *scores* than about either content *standards* or the public *reporting* of test results, all three having been described as drivers of educational reform (NRC, 1999; Mabry, Poole, Redmond & Schultz, 2003; Marion & Gong, 2003). Teachers' test-related anxiety infringed on their job satisfaction and drove a surprisingly high number of participants from the fourth-grade during the two years of this study.

Teachers' deepest concerns about the impact of current accountability initiatives, as indicated by both frequency and poignancy of expression, centered on their students. As attention to test scores rose, teachers reported that children were increasingly suffering from test stress, a few students in one school requiring therapeutic intervention. Their efforts to help students cope with the pressure, teachers said, were essentially futile. While most administrators extolled data-driven improvements, one principal summarized the worry:

We're not teaching robots.... I guess I don't know if No Child Left Behind is really talking about a child or a piece of data that isn't being left behind"
(Highland principal, June 28, 2005).

References

- Anderson, P. (2005, May 25). NCLB is not a “mandate,” says GAO in clarification [letter to the editor]. *Education Week*, p. 33.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, D. C.: Authors.
- Archer, J. (2005a, April 13). Connecticut pledges first state legal challenge to NCLB law. *Education Week*, p. 38.
- Archer, J. (2005b, May 25). States eyeing expense of hand-scored tests in light of NCLB rules. *Education Week*, p. 1.
- Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). Standards for educational accountability systems. *CRESSST Line*, Winter, 1–4.
- Beach, K. D., Gee, J. P., Greeno, J. G., Haertel, E. H., Lee, C. D., Mehan, H. B., Mislavy, R. J., Moss, P. A., & Pullin, D. C. (2003, April). *Opportunity to learn and assessment: Psychometric and sociocultural perspectives*. Presentation to the American Educational Research Association, Chicago.
- Behuniak, P., DeVito, P., Rivera, C., & Fremer, J. (2001, June). *In search of fair large-scale assessments*. Presentation to the CCSSO Large-Scale Assessment Conference, Houston.
- Benson, D., Moses, M., Saul, D., Erb, J., Otto, D., Phillips, D., & Erickson, J. (2004). *No Child Left Behind fails the tests of common sense*. Western States Benchmarking Consortium, unpublished memorandum.
- Bergeson, T. (2005). *New Year's address to Washington educators* [video]. Olympia, WA: Office of the Superintendent of Public Instruction.
- Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18–20.
- Berliner, D.C. & Biddle, B.J. (1995). *The manufactured crisis: Myths, fraud and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231–268.
- Borko, H., Wolf, S.A., Simone, G., & Uchiyama, K.P. (2003). Schools in transition: Reform efforts and school capacity in Washington state. *Educational Evaluation and Policy Analysis*, 25(2), 171–201.
- Broder, D. S. (2005, February 28). The unsustainable budget. *Oregonian*, p. D7.

- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Burlingame, M. (1993). Compliance and local educational politics and policy. *Planning and Changing*, 24(1.2), 86–100.
- Cannell, J. J. (1987). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice*, 7(2), 5–9.
- Chatterji, M. (2005). Evidence on "what works": An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher*, 33(9), 3–13.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142.
- Cook, H. G., Fitton, R., Viator, K., Phillips, S. E., & Gertz, S. (2001, June). *Politics and high-stakes testing*. Presentation to the CCSSO Large-Scale Assessment Conference, Houston.
- Cuban, L. (1988). Why do some reforms persist? *Educational Administration Quarterly*, 24(3), 329–35.
- Darling-Hammond, L. (1990). Instructional policy into practice: "The power of the bottom over the top." *Educational Evaluation and Policy Analysis*, 12(3), 339–347.
- Davis, M. R. (2005, February 9). Utah is unlikely fly in Bush's school ointment. *Education Week*, pp. 1, 21.
- Davis, M. R., & Hoff, D. J. (2005, April 20). Questions linger over NCLB policy shifts. *Education Week*, pp. 21–22.
- Denzin, N. K. & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher*, 32(7), 31–38.
- Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. NY: Macmillan.
- Elmore, R., Fuhrman, S., Koretz, D., Herman, J. L., Linn, R. L., & Baker, E. L. (2001, April). *Holding accountability systems accountable: Research-based standards*. Presentation to the annual meeting of the American Educational Research Association, Seattle, WA.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York: Macmillan.
- Fontana, A. & Frey, J. H. (1994). Interviewing: The art of science. In Denzin, N. K. & Lincoln, Y. S. (Eds.), *Handbook of qualitative research* (pp. 361–376). Thousand Oaks, CA: Sage.

- Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27–32.
- Fuhrman, S. H. & Elmore, R. F. (1990). Understanding local control in the wake of state education reform. *Educational Evaluation and Policy Analysis*, 12(1), 82–96.
- Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Galley, M. (2003, June 18). More errors are seen in the scoring of tests, Boston researchers say. *Education Week*, 22(41), 10.
- Gewertz, C. (2005, April 20). Critics question use of offshore firms for online tutoring. *Education Week*, p. 14.
- Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity*. Buckingham, England: Open University Press.
- Glaser, B. G. & Strauss, A. I. (1967). *The discovery of grounded theory*. Chicago, IL: Aldine.
- Goodson, I. & Foote, M. (2001). Testing times: A school case study. *Education Policy Analysis Archives*, 9(2). Retrieved July 23, 2006, from <http://epaa.asu.edu/epaa/v9n2/>.
- Green, S. K., Alderman, G., & Liechty, A. (2004). Peer tutoring, individualized intervention, and program monitoring with at-risk second-grade readers. *Preventing School Failure*, 49(1), 11–18.
- Haertel, E. H. (2002, April). *Technical considerations in the use of NAEP to confirm states' achievement gains*. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2–7.
- Hall, B., McDonald, J., Scherich, H., Vickers, D., & Zebrowski, M. (2001, June). *Performance scoring—The challenges and opportunities in today's assessment world*. Panel presentation to the CCSSO Large-Scale Assessment Conference, Houston.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved July 23, 2006, from <http://epaa.asu.edu/epaa/v8n41/>.
- Harmon, M. H. (1995). *Responsibility as paradox: A critique of rational discourse on government*. Thousand Oaks, CA: Sage.
- Hendrie, C. (2005a, May 4). NCLB cases face hurdles in the courts. *Education Week*, pp. 1, 22.
- Hendrie, C. (2005b, April 20). NCLB transfer policy seen as flawed. *Education Week*, pp. 1, 14–15.

- Herman, J. L., Baker, E. L., & Linn, R. L. (2004, Spring). Accountability systems in support of student learning: Moving to the next generation. *CRESST Line*, 1–7.
- Hill, R. K. (2002, April). *Examining the reliability of accountability systems*. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- Hill, R. K., & DePascale, C. A. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice*, 22(3), 12–20.
- Hoff, D. J. (2005a, May 4). Education department fines Texas for NCLB violation. *Education Week*, pp. 27, 31.
- Hoff, D. J. (2005b, April 13). States to get new options on NCLB law. *Education Week*, pp. 1–38.
- Hoff, D. J. (2005c, March 9). Texas stands behind own testing rule; Move puts state at odds with NCLB policy. *Education Week*, pp. 1, 23.
- Keller, B. & Sack, J. L. (2005, April 27). Union, states wage frontal attack on NCLB. *Education Week*, pp. 1, 18.
- Kerr, H. V. & Verhaeghe, J. P. (2005). Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *Journal of Experimental Education*, 73(4), 291–329.
- Knapp, M. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research*, 67, 227–266.
- Koretz, D. (2001, April). *Toward a framework for evaluating gains on high-stakes tests*. Paper presentation to the annual meeting of the National Council on Measurement in Education, Seattle, WA.
- Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.
- LeCompte, M. D. & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego: Academic Press.
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797–832.
- Lemire, C. (1998, May 14). Court rejects TAAS score dispute. *Austin American-Statesman*.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.

- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3–16.
- Louis, K. S., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27(2), 177–204.
- Mabry, L., Poole, J., Redmond, L., & Schultz, A. (2003). Local impact of state-mandated testing. *Education Policy Analysis Archives*, 11(22).
- Marion, S. & Gong, B. (2003, October). *Evaluating the validity of state accountability systems*. Paper presented to the annual Reidy Interactive Lecture Series, Nashua, NH.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3), 279–300.
- McLaughlin, M. S. (1991). Test-based accountability as a reform strategy. *Phi Delta Kappan*, 73(3), 248–252.
- Meier, T. (1995). Standardized tests: A clear and present danger. In Levine, D., Lowe, R., Peterson, B., & Tenorio, R. (Eds.) *Rethinking Schools: An Agenda for Change*. New York: The New Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education, Macmillan.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Miskel, C. & Song, M. (2004). Passing *Reading First*: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, 25(2), 89–109.
- National Research Council (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- No Child Left Behind Act. (2001). Public Law No. 107–110, 115 Stat. 1425, 2002.
- Olson, L. (2004, September 22). No Child Left Behind Act changes weighed. *Education Week*, pp. 31, 34.
- Olson, L. (2005a, May 25). Florida gains flexibility on NCLB provisions. *Education Week*, pp. 18, 20.
- Olson, L. (2005b, April 6, 2005). NCLB study: State support for schools' improvement varies. *Education Week*, p. 9.

- Olson, L. (2005c, July 13, 2005). Requests win more leeway under NCLB. *Education Week*, pp. 1, 20.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Pellegrino, J. W., & Goldman, S. R. (2002). Be careful what you wish for—you may get it: Educational research in the spotlight. *Educational Researcher*, 31(8), 15–17.
- Raywid, M. A. (1990). Pseudo-reform, incremental reform, and restructuring. *Phi Delta Kappan*, 72, 139–143.
- Rossman, G. B., Corbett, H. D. & Dawson, J. A. (1984). Intentions and impacts: A comparison of sources of influences on local school systems. *Urban Education*, 21(1), 86–106.
- Rotherham, A. J. (2005, January 12). In education, no pundit left behind. *Oregonian*, p. B9.
- Rubin, H. J. & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, 42 (1), 3–42.
- Sack, J. L. (2005, April 27). Utah passes bill to trump 'No Child' law. *Education Week*, pp. 22–25.
- Sarason, S. B. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston: Allyn and Bacon, Inc.
- Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it's too late?* San Francisco: Jossey-Bass.
- Schaffer, E. C., Nesselrodt, P. S., & Stringfield, S. C. (1997). *Impediments to reform: An analysis of destabilizing issues in ten promising programs*. Arlington, VA.: Educational Research Service.
- Shepard, L. A. (2002, April). *Building bridges between classroom and large-scale assessments*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Shipp, D. (2003). Pulling together: Civic capacity and urban school reform. *American Educational Research Journal*, 40(4), 841–878.
- Sipple, J. W., Killeen, K., & Monk, D. H. (2004). Adoption and adaptation: School district responses to state imposed learning and graduation requirements. *Educational Evaluation and Policy Analysis*, 26(2), 143–168.
- Sirotnik, K. A. (Ed.) (2004). *Holding accountability accountable: What ought to matter in public education*. New York: Teachers College Press.

- Skinner, R. A. (2005, January 6). The state of the states. *Quality Counts 2005: No Small Change*. Bethesda, MD: *Education Week*.
- Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28, 521–542.
- St. Pierre, E. A. (2002). "Science" rejects postmodernism. *Educational Researcher*, 31(8), 25–27.
- Sternberg, R. J. (2002). The "Janus principle" in psychometric testing: The example of the upcoming SAT-I. *The Score*, newsletter of American Psychological Association Division 5, Evaluation, Measurement, and Statistics, 24(2), 3–5.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Strauss, A. & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Thousand Oaks, CA: Sage.
- Stringfield, S. C., & Yakimowski-Srebniak, M. E. (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City Public Schools. *American Educational Research Journal*, 42(1), 43–75.
- Toenjes, L. A. & Dworkin, A. G. (2002). Are increasing test scores in Texas really a myth, or is Haney's myth a myth? *Educational Policy Analysis Archives*, 10(17), 1–11.
- Tyack, D. & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Tyack, D. & Tobin, W. (1994). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31, 453–479.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press.
- Washington State Senate. (1992). *Washington Substitute Senate Bill 5953: Act relating to education*. Olympia, WA: Authors.
- Washington Education Association. (2004). *Washington Education Association testimony to the House Education Committee*. Retrieved December 5, 2005, from http://www.washingtonea.org/static_content/ed_reform/testimony/Hse9_20_04.htm.
- Wise, A. (1979). *Legislated learning: The bureaucratization of the American classroom*. Berkeley: University of California Press.
- Wolcott, H. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.

Wood, S. O. (2005, September 1). Washington outpaces Oregon on SAT—again. *Oregonian*, pp. B1, B3.

Zehr, M. A. (2004, September 8). Wary districts shift or forgo federal funds. *Education Week*, p. 6.

About the Authors

Linda Mabry

Washington State University Vancouver

Jason Margolis

Washington State University Vancouver

Email: mabryl@vancouver.wsu.edu

Linda Mabry's research interests relate to the assessment of the achievement of K–12 students and to program evaluation.

Jason Margolis's research interests include educational reform and teacher development.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermadorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Robert Bickel

Marshall University

Casey Cobb

University of Connecticut

Gunapala Edirisooriya

Youngstown State University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

David C. Berliner

Arizona State University

Gregory Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Mark E. Fetler

California Commission on
Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Ohio University

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Cannata
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Públicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España