

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. EPAA is a project of the [Education Policy Studies Laboratory](#).

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the [ERIC Clearinghouse on Assessment and Evaluation](#) and are permanently archived in *Resources in Education*.

Volume 11 Number 22

July 17, 2003

ISSN 1068-2341

Local Impact of State Testing in Southwest Washington

Linda Mabry

Jayne Poole

Linda Redmond

Angelia Schultz

Washington State University Vancouver

Citation: Mabry, L., Poole, J., Redmond, L., Schultz, A. (July 18, 2003). Local impact of state testing in southwest Washington. *Education Policy Analysis Archives*, 11(21). Retrieved [date] from <http://epaa.asu.edu/epaa/v11n22/>.

Abstract

A decade after implementation of a state testing and accountability mandate, teachers' practices and perspectives regarding their classroom assessments and their state's assessments of student achievement were documented in a study of 31 teachers in southwest Washington state. Against a background of national trends and standards of psychometric quality, the data were analyzed for teachers' beliefs and practices regarding classroom assessment and also regarding state assessment, commonalities and differences among teachers who taught at grade levels tested by the state and those who did not, teachers' views about the impact of state assessment on their students and their classrooms, and their views about whether state testing promoted educational improvement or reform as intended. Data registered (1) teachers' preferences for

multiple measures and their objections to single-shot high-stakes testing as insufficiently informative, unlikely to promote valid inferences of student achievement, and often distortive of curriculum and pedagogy; (2) teachers' objections to the state test as inappropriate for nonproficient speakers of English, for students eligible for special services, and for impoverished students; and (3) teachers' preferences for personalized assessments respectful of student circumstances and readiness, rather than standardized assessments. Teachers' practical wisdom thus appeared more congruent than the state testing program with measurement principles regarding (1) multiple methods and (2) validation for specific test usage, including usage with disadvantaged subgroups of test-takers. Findings contrasted a distinction of emphasis: state focus on "*testing* students" as distinct from teachers' focus on "*testing students*."

By 2001-02, standards and standards-based testing were being implemented in 49 states to evaluate school and student performance (Meyer, Orlofsky, Skinner & Spicer, 2002, p. 74), all save Iowa where the state requires district standards (Neuman, 2002) and where it has been reported that virtually all school districts administer the Iowa Test of Basic Skills (ITBS) (Bond, Braskamp, van der Ploeg & Roeber, 1996; Mabry & Daytner, 1997). Formal purposes for standards-based state testing programs typically include statements of intent to improve student learning. For example, Substitute Senate Bill 5953 (SSB 5953), the origin of Washington state's current testing program, opens with these words:

If young people are to prosper in our democracy and if our nation is to grow economically, it is imperative that the overall level of learning achieved by students be significantly increased. To achieve this higher level of learning, the legislature finds that the state of Washington needs to develop a performance-based school system. . . . [T]he state needs to hold schools accountable for their performance based on what their students learn. . . . [I]t will be necessary to set high expectations for all students, to identify what is expected of all students, and to develop a rigorous academic assessment system to determine if these expectations have been achieved. (Washington State Senate, 1992, pp. 1-2)

A decade after implementation of this legislation, has Washington's accountability plan had the intended effect? Have the state content standards, the Essential Academic Learning Requirements (EALRs), and the standards-based test, the Washington Assessment of Student Learning (WASL), improved student learning? With awareness that there have been few empirical studies of the effects of the standards movement nationally (Swanson & Stevenson, 2002) and with particular interest in the local context, an interview study of 31 teachers was undertaken in 2001-02 to discover the impact of reform-oriented, standards-based state testing in southwest Washington, with emphasis on whether it had encouraged changes in classroom practices which promoted improved learning.

Context

National education reform

The implicit theory of action underlying test-driven accountability systems is that testing will improve student learning through provision of accurate data supporting valid interpretations of student achievement, with scores used to identify those who will receive rewards and sanctions, ultimately motivating improved teaching and learning (Baker, 2002). This theory implies that teachers and students are extrinsically motivated, that test scores and the rewards and sanctions they trigger are motivating in the manner intended, and that teachers and students are not working as hard as they could and should (Elmore, 2002). When the nodes in this chain of logic are examined in sequence (see Figure 1), it becomes clear that threats to any link in the chain can result in testing that not only does *not* improve learning but may even be counterproductive.

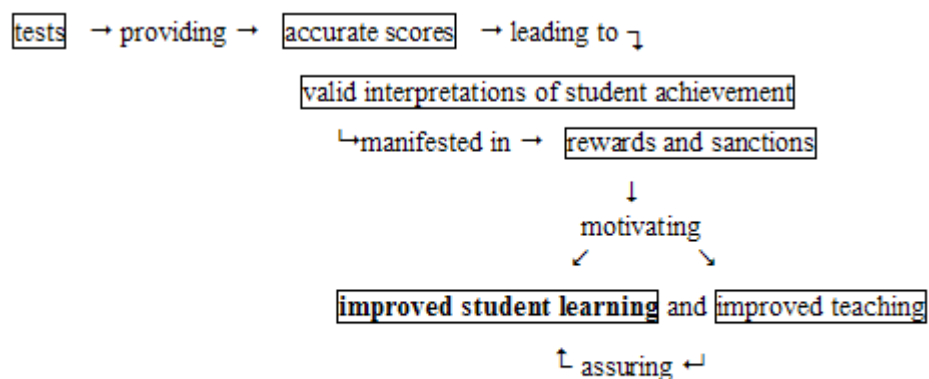


Figure 1. Theory of action in high-stakes state testing and accountability, graphic based on presentations by Baker, 2002 and Elmore, 2002

For example, what if test scores do *not* provide accurate data but if, as has sometimes been charged, the tests are biased against racial and ethnic minorities, females, or the poor? What if rewards and sanctions do *not* motivate teachers to improve teaching but, rather, motivate them to subvert and distort their practice through teaching to the test or "multiple-choice teaching" (Smith, 1991, p. 10)? While the theory of action suggests the mechanisms and sequencing through which testing can improve teaching and learning, it simultaneously suggests the critical junctures at which testing can undermine teaching and learning.

In high-stakes testing, theoretical implications matter much less than real-life implications. Empirical data indicate that scores do tend to rise in the years following the implementation of a new test (Linn, 2000), consistent with the theory of action. Washington state's test data also exhibits this trend (see Table 1), although not uniformly. But whether the higher scores reflect increased student learning is unclear (Haladyna, Nolen & Haas, 1991; Mabry, Aldarondo & Daytner, 1999; Shepard & Smith, 1988; Smith & Rottenberg, 1991). Are the scores accurate, and are they triggering appropriate consequences that yield improved teaching and learning?

Table 1 Trends in Washington state test scores, 1997-2001: Percentages

**of students meeting state standards in reading, math, and writing,
based on data available online at www.k12.wa.us**

Test subjects and grades	Scores by years				
	1996-97	1997-98	1998-99	1999-2000	2000-01
Reading					
grade 4	47.9	55.6	59.1	65.8	66.1
grade 7		38.4	40.8	41.5	39.8
grade 10			51.4	59.8	62.4
Mathematics					
grade 4	21.4	31.2	37.3	41.8	43.4
grade 7		20.1	24.2	28.2	27.4
grade 10			33.0	35.0	38.9
Writing					
grade 4	42.8	36.7	32.6	39.4	43.3
grade 7		31.3	37.1	42.6	48.5
grade 10			41.1	31.7	46.9

Bar graph based on *fourth grade reading scores* (first row in the table above), rounded to the nearest whole number, to visualize score increases more clearly

Scores					
68					
67				65.8	66.1
66					
65					
64					
63					
62					
61					
60			59.1		
59					
58					
57		55.6			
56					
55					
54					
53					
52					

51					
50					
49	47.9				
48					
47					
46					
Years	1996-97	1997-98	1998-99	1999-2000	2000-01

The consequences of state testing in the U.S., where the stakes are high and getting higher, indicate widespread acceptance—at least implicitly—of the theory of action. Currently, 43 states require school report cards (including Washington), with two more in development, and 20 of these require that the report cards be sent home to parents. Twenty states (not including Washington) have the authority to impose serious sanctions on low-performing schools: school closure or reconstitution, student transfers, and loss of funding; three more states will be able to do so within two years. Eighteen states (not including Washington) provide rewards to high-performing or improved schools, with two more set to do so within two years. Fifteen states use test scores alone, with no additional evidence, to evaluate schools (Meyer, Orlofsky, Skinner & Spicer, 2002).

The difficulty charter schools are experiencing in trying to raise test scores (e.g., Gewertz, 2002) is heightening awareness that raising test scores in straightened educational circumstances is not easy. Perhaps because of this, test-triggered stakes are increasingly being borne by students who are relatively defenseless (Elmore, 2002). In particular, in seventeen states, a number that will increase by seven in the next two years, adolescents cannot graduate from high school without passing exit or end-of-course exams (Washington will require a graduation test in 2008). An elementary or middle school child's promotion to the next grade is contingent on test scores in four states (not including Washington), a number that will double in the next two years. Remediation is required for students failing promotion, end-of-course, or high school graduation exams in seventeen states, most but not all of which provide funds for the remedial instruction (Meyer, Orlofsky, Skinner & Spicer, 2002).

The newly reauthorized Elementary and Secondary Education Act (2001), dubbed "no child left behind" (NCLB) and sometimes derisively called "no child left untested," furthers the trend toward more state testing and higher stakes. Stakes include federal Title 1 funding and now, for underachieving schools, requirements to provide school choice to parents in year 2 of a school's continuing low test scores, tutoring with parental choice as to providers in year 3, replacement of curriculum and/or staff in year 4, and reconstitution in year 5. The basis of these sanctions is state test scores. In Washington, few schools currently meet NCLB standards: only 36 of 1162 elementary schools, 19 of 554 middle schools, and 13 of 505 high schools (*Oregonian*, 2002).

Superseding the Goals 2000 call for a national system of tests in 1994, (Note 1) NCLB requires increased state testing, including standards-based assessments of reading and math for all students in grades 3-8. In order to receive Title 1 funds, the law requires attainment of proficiency by all students—including minorities, students with limited proficiency in English, and low SES students—within twelve

years and proportional annual yearly progress (AYP) in the interim. To discourage states from using easy tests that might distort achievement or lower expectations, the law also requires that scores on state tests be confirmed against scores on the National Assessment of Educational Progress (NAEP). (Note 2)

The AYP targets are about double the score increases empirically documented by NAEP over time, which suggest that it might take not twelve but more than 100 years, by optimistic estimate, to reach the required 100% proficiency. The AYP targets have been judged especially "unrealistic" for schools and districts where small enrollments of disadvantaged subgroups of students will result in statistically unstable results (Haertel, 2002). The targets also appear painfully unrealistic for chronically under-resourced urban schools (Lewis, 2002; Yakimowski, 2002). National policy thus exhibits confidence in a theory of action that is empirically suspect.

State policy

The standards-based, test-driven educational reform initiative mandated by the legislature in SSB 5953 in 1992 lists four purposes for the state of Washington's accountability system:

- to assess students' academic learning
- to evaluate instructional practices
- to select students for remediation
- to hold schools accountable for student learning (Washington State Senate, 1992, p. 10).

These are very similar to the four purposes for assessments recently listed by Shepard (2002)—diagnosis, monitoring, student selection, and program evaluation—with the warning that making a test more valid for one purpose might make it less valid for a competing purpose. Frequent similar admonitions from the measurement and evaluation communities indicate that multiple purposes for a single test are usually problematic, as different purposes are often in unwitting conflict, undermining achievement of any of the goals (e.g., Mabry, 1999). For example, tests used for school accountability have often proved vulnerable to "score pollution" as school personnel administering the tests succumb to pressure to raise scores through a variety of means, some ethically, legally, or statistically questionable (Haladyna, Nolen & Haas, 1991; Haney, 2000; Linn, 2000; Sternberg, 2002).

Score increases are not always credible, as evidenced by discrepancies between some state NAEP scores and scores on the state test (e.g., Haney, 2000) and by the so-called Lake Wobegon effect—states' insistence that more than half of their students were "above average" (Cannell, 1987), a statistical impossibility. As educators scramble to raise scores to protect their schools, students, and themselves from high-stakes penalties, improved state test scores may not necessarily reflect improved student achievement. Inflated scores would obstruct understanding of students' academic learning and would obstruct identification of students needing remedial assistance—two goals of Washington state's accountability system.

The Washington Assessment of Student Learning (WASL) tests literacy and math at grades 4, 7, and 10 and offers multiple-choice and constructed-response items, both short and extended writings. Described as a criterion-referenced assessment aligned to state standards (Meyer, Orlofsky, Skinner & Spicer, 2002, p. 75), the WASL is administered in late Spring. Student performance is judged to be "above standard," "meets standard" (the required level of proficiency), "below standard," and "well below standard." In 2001, schools and districts were required to reduce by 25% the number of students not meeting the state's required standard and to include in public reporting their goals and plans to do so (online at the state education agency's website, www.k12.wa.us). As noted, in comparison to some states, the stakes associated with the WASL are relatively low: schools are not threatened with closure or reconstitution; funds are not withheld because of low scores; students in grades 4 and 7 are not retained at grade level or compelled into remedial education if they do not meet standards; high school students' eligibility for graduation will not be contingent upon WASL scores until 2008 (Note 3) (Meyer, Orlofsky, Skinner & Spicer, 2002, pp. 74-75).

Is Washington's testing program having the intended effect, assuring that "the overall level of learning achieved by students be significantly increased"? State statistics generally suggest improved achievement (see Table 1) but, as of 2001, national statistics indicated that less than a third of Washington's fourth- or seventh- graders had scored at the "proficient" level on NAEP reading, writing, math, or science tests (Orlofsky & Olson, 2001). Of course, it might be that the scores reflect state learning goals but not national learning goals. It might also be that the state's standards-based testing program is improving learning but not yet measurably since, elsewhere, indications have been found that state reforms are resulting in teachers' adoption of classroom practices consistent with standards (Swanson & Stevenson, 2002). Local evidence of teachers' acceptance of Washington's state standards and of positively evolving classroom practices, if occurring, might suggest gradual improvement which could become measurable in the future.

The research reported here investigated the resonance between state testing and classroom assessments, whether feedback from the WASL helps teachers understand their students' achievements and plan more effective learning opportunities, whether local classroom practices are changing, whether state testing is encouraging the alignment of curriculum to state standards and, if so, whether the alignment is educationally beneficial.

Method

The approach to the study undertaken in Fall 2001 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 and methods respectful of emergent design, multiple perspectives, and inductive analysis (Denzin, 1989, 1997; Denzin & Lincoln, 1994; Erickson, 1986; Mabry, 2002; Merriam, 1998; Stake, 1978; Wolcott, 1994). Two data collection methods were employed: review of documents (Hodder, 1994) related to testing in Washington state and, more importantly, semi-structured interviews (Fontana & Frey, 1994; Rubin & Rubin, 1995) of practicing teachers in

the local area.

After approval of the study by a university Institutional Review Board and signed consent from each interviewee, graduate students at Washington State University Vancouver (Note 4) interviewed 31 local teachers in Fall 2001. The sampling strategy was purposeful rather than representative or randomized, with each graduate student identifying and interviewing two teachers who taught a subject at a grade level of specific interest to the interviewer. (Note 5) This subject selection strategy maximized the sensitivity of the interviewers to each teacher's subject area and grade level.

Of the 31 teachers interviewed, 19 taught in high schools, 5 in middle schools, and 7 in elementary schools (see Table 2). Their teaching experience totaled 547 years, with an average of 18 years each and a range of 1-40 years. Nineteen interviewees were female and 12 were male. All of the teachers' schools were located in southwest Washington state, and all but one of these was a public school. The teachers included 13 who taught subject areas and grade levels tested by the WASL and 18 who did not. Of the teachers whose students were tested on the WASL, 9 taught in high schools, 2 in middle schools, and 2 in elementary schools.

Table 2
Teachers interviewed, the subjects and grade levels they taught,
and whether these subjects and grades levels were tested using
the WASL (n = 31)

Level	Subject/grade	Tested subject at this grade level?	Teachers (by pseudonym) of this subject at this grade level and years of experience
High school <i>n=19,</i> <i>9 in tested grades</i>	English-language arts*	YES	Ms. Apple, 3 years Ms. Brush, 15 years Mr. Carr, 7 years Mr. Dustin, 20 years Ms. Hand, 22 years Ms. Kroner, 7 years Mr. Twain, 25 years Ms. Underwood, 20 years
	mathematics**	YES	Mr. Alder, 19 years
	science	no	Mr. Liu (biology), 17 years Mr. Ming (biology), 17 years Mr. Ochre (biology), 9 years

			Ms. Vargas, 20 years Ms. Walker, 20 years Mr. Banks, 34 years
	family and consumer ed	no	Ms. Crane, 30 years Ms. Doe, 14 years
	foreign language	no	Ms. Good, 22 years
	social studies	no	Mr. Inder, 1 year
Middle school	English-language arts	YES	Ms. Frank, 12 years Ms. Nunn, 5 years
<i>n=5, 2 in tested grades</i>	history	no	Mr. Eggle, 25 years
	social studies and other	no	Ms. Grant, 18 years
	grade 6	no (private school)	Ms. Smith, 16 years
Elementary school	grade 1	no	Ms. Park, 16 years
	grade 2	no	Ms. Hallo, 13 years
<i>n=7, 2 in tested grades</i>	grade 3	no	Ms. Jones, 30 years Ms. Quinn, 40 years
	grade 4	YES	Ms. Roberts, 21 years Mr. Exeter, 8 years
	grade 5	no	Mr. Felix, 21 years

* One teacher taught English-language arts and also civics and philosophy.

** This teacher taught math and also P.E.

A collaboratively constructed interview protocol (see Exhibit 1) guided semi-structured interviewing (Fontana & Frey, 1994; Rubin & Rubin, 1995). Interviews lasted approximately 45 minutes each. Interviewers attempted to capture

as many direct quotations as possible, with some interviews tape-recorded with permission of the interviewees and others recorded in hand-written notes typed up soon thereafter. For purposes of developing a high-quality database with strong internal validity (Campbell & Stanley, 1963) or descriptive validity (Maxwell, 1992), a comprehensive validation strategy (Mabry, 1998) was used, with each interview written up and presented to the interviewee with a request for review, correction, and elaboration.

Exhibit 1. Protocol for semi-structured interviewing of teachers

- How many years have you been teaching? What grade levels and subject areas have you taught? Has all of your teaching occurred in the state of Washington?
- How do you assess your students' achievement?
- How did you develop your approach to student assessment? Why did you take this approach? What influenced your thinking? How long did it take to develop? How has it evolved over time (if it has)?
- Have you had training in assessment? If so, how much training have you had? How would you describe the type of training you have had? Has your assessment training been related to specific content areas?
- As the state has developed requirements for student learning and for assessing student achievement, has your teaching changed? If so, what has changed about your teaching? Do you consider the changes to be improvements?
- How do you feel about the WASL (Washington Assessment of Student Learning)? Why?
- How do you prepare your students for your assessments (if you do)? How do you prepare them for state assessments (if you do)?
- If you were to change your classroom assessments, what would you like to do differently? If state assessments were to change, what type of change would you favor?
- Does your school or district require testing (other than state testing)? Are tests part of your school's or district's graduation requirements for high school students?
- Is there anything you would like to add?

Thank you very much for your time and information! I will type up my notes from this interview and give them to you. I would very much appreciate it if you would read the notes and make any corrections to improve accuracy. If there is anything you would to add at that time, I hope you will feel

free to make additions then. Again, *many thanks!*

Data analysis was emergent in character, with meaning sought in the data without reference to *a priori* categories (Denzin & Lincoln, 1994; Erickson, 1986; Mabry, 2002; Wolcott, 1994). Analysis involved four phases and two validation efforts. In the first phase, pairs of graduate students analyzed their four interviews for patterns, including commonalities and distinctiveness across their four subjects. This thematic content analysis (LeCompte & Preissle, 1993; Miles & Huberman, 1994) and the resulting preliminary interpretations were written up in eight separate preliminary reports. In the next phase, the first author conducted a similar content analysis across the eight student reports, identifying 29 themes overall and grouping them in four emergent categories: (1) classroom impact, (2) student impact, (3) teacher impact, and (4) teachers' perspectives (see Table 3).

Table 3
Themes emerging from content analysis of teacher interview data, identified from eight preliminary interview reports and grouped into four categories

	Themes	Interview reports							
		1	2	3	4	5	6	7	8
A	Classroom impact								
	Teachers' approaches to classroom assessment	X	X	X	X	X	X	X	X
	Training in assessment for teachers	X	X	X	X	X	X	X	X
	Changes in assessments over time		X			X		X	
	Usefulness of the WASL for classroom practice					X			
	Impact of state standards/tests on curriculum and instruction (or resistance to impact)	X	X	X	X	X	X	X	X
	Preparation in class for the state test		X	X	X	X	X	X	
	Impact of the WASL on classroom assessments			X			X		
	Impact of the WASL on classroom environment					X			
B	Student impact								
	Student accountability based on WASL scores (e.g., graduation, retention)	X		X			X		X
	Equity to students, including						X		
	• Students whose first language is not English			X	X				
	• Special education students			X	X				

	• Transfer students				X			
	• Minority students			X	X			X
	• Students in difficult circumstances (including SES)			X				
	Impact of the WASL on students' self-esteem/stress/anxiety		X	X		X		X
C	Teacher impact							
	School/teacher accountability based on WASL		X	X	X	X	X	X
	Pressure to perform well on the WASL			X	X	X		
	Impact of the WASL on teacher professionalism			X				
	Contrasts of interest to us, including							
	• Public and private schools					X		
	• Tested and non-tested grades/subject areas by the WASL	X		X	X	X	X	X
	• Teacher assessments and state assessments	X	X	X	X	X	X	X
D	Teachers' perspectives							
	Teacher approval or What teachers like about state testing		X			X	X	X
	Teacher disapproval or What teachers do not like (or would change) about state testing		X	X		X	X	X
	Questioning the constructs tested (or that should be tested) by the WASL	X	X	X		X	X	X
	Questioning the difficulty level of the WASL				X	X		
	Scoring concerns		X				X	X
	Questioning the expense of testing							X

(Note. An **X** indicates that data related to the theme (listed by row) were found in the preliminary report (listed by column) in phase 2 of data analysis. Further review in phase 3 identified additional sources of data on these themes, revisions to these themes, and additional themes.)

A third data analysis phase involved micro-review by the authors of the entire data set for comprehensive identification of all data points related to each theme and category. The final phase of analysis was the identification, drafting, and formalization of findings. A draft of the resulting manuscript was offered for review

and critique to all 16 interviewers in a second validation effort.

The data and findings were structured for reporting according to the four major thematic categories. The teachers quoted are identified by pseudonyms.

Classroom impact

Consistent with other findings about the impact of standards-based reform (Swanson & Stevenson, 2002), the data made clear that all of the teachers interviewed were highly aware of the state reform initiative and that state policy was definitely felt in local schools and classrooms. Veteran teachers stood as witnesses to changes ushered in by the reform efforts. For example, a teacher with forty years' experience observed, "Early in my career, there was very little emphasis on assessment. This has changed, most recently because of the state Essential Learnings" (Ms. Quinn). She, among others, indicated that she had seen public school assessments evolve over the years from reliance on intuitive teacher judgments to formal state standards, the Essential Academic Learning Requirements or EALRs.

Response to state initiative

In response to Washington's state standards and the state test, the WASL, most interviewed teachers had adjusted their classroom practice, they reported, some more than others. While most teachers said their instructional styles had not changed, many said the content they taught had altered, consistent with other research indicating that teachers feel state frameworks are redefining curricula (Shore, 2002). Some teachers reported positively that new state standards provided explicit objectives which helped focus their teaching. For example, one said, "It gives students and teachers a target" (Ms. Frank). Another commented, "If you are going to try to do the job right, you should always have [the EALRs] beside you so you can see what you're doing is on target" (Ms. Good). These teachers had approvingly accepted the EALRs as their teaching goals. A high school science teacher suggested that acceptance might sometimes have been compelled rather than willing:

In general, [the WASL] is a good idea because it has forced people to be accountable. Many kids from middle school didn't have the basics, and we had to spend time re-teaching what they should already have known. (Ms. Vargas)

Some teachers expressed ambivalence or frustration regarding the superimposing of state goals over their own aims and approaches. For example, a high school teacher expressed hope that aligning her teaching to the EALRs had "tightened up" her teaching but also said she found it frustrating when this forced elimination of her successful hand-crafted units: "It's hard to ditch your pet projects" (Ms. Good). A third grade teacher who had previously taught thematically said she now taught subject-by-subject, with special attention to state content standards, a change she described not as an improvement but as "a necessity in the constantly changing world of education" (Ms. White).

With the state graduation test postponed until 2008, classroom practices were less

affected at the high school level, according to the teachers, and less evident in content areas which were not yet tested by the WASL. Even so, many high school teachers indicated strong impact of the test on their practices. For example, the only interviewee who baldly admitted to teaching to the state test was a high school teacher who said he had been directed to do so by his principal because there was “a lot at stake” (Mr. Ochre). Greater impact was apparent at the elementary level, particularly in fourth grade, a tested grade. At this grade level, wholesale displacement of the curriculum was noted by some, including one fourth grade teacher who said, “Teachers in [my] building spend from about November to mid-April focused on the WASL” (Ms. Roberts).

Classroom assessments

Every teacher reported using a variety of assessment methods and techniques in the classroom. Often, these featured performance assessment, their reported practices ranging from observations of student performances to portfolios and projects. Variations among the teachers' assessment ideologies and practices suggested adaptations harmonious with personal style, the range surprising some interviewers. For example, Ms. Hand and Ms. Kroner were described by their interviewer as having “vastly different takes on what constituted appropriate assessment, yet both were excellent teachers who were obviously very dedicated to their profession.” Even teachers who taught the same subjects and grade levels approached assessment differently (e.g., Mr. Twain and Ms. Underwood, high school English; Ms. Vargas and Ms. Walker, high school science). Suggesting adaptations based on experience and changes in student populations, teachers consistently described continuous efforts across time as “constantly evolving each year as my class changes and the world around them changes” (Mr. Banks).

Using assessment to ensure student success, rather than to identify weaknesses for remediation or penalty, emerged as an important distinction between classroom and state assessments in comments from some teachers, such as:

I want kids to be successful in my classroom. I'm not there to fail students. I'm there to teach students. [For] those with low academic abilities, if you put too much emphasis on testing, you would see a high failure rate. (Mr. Liu)

Another who offered an earnest rationale for using assessment to improve achievement rather than to punish students said:

Over my 17 years of teaching, I've really changed my approach to student assessment. Initially, I started out really being worried about content. The value of their grade was based more on testing—maybe 80%, 90%—less on what they did in the classroom, less on behavior. Over the 17 years, I've changed that. Maybe it's not so important *what* they're learning but *how* they're going about doing it, how they're approaching what they're doing in class. I've shifted my emphasis from content and techniques to behavior and work-related skills. . . .

I want kids to be successful in the classroom. If I based [grades] strictly on content, I'd see too many kids failing. I think today more kids are

coming to my classroom without the tools needed to be successful in terms of learning the same level of content that I expected 17 years ago. So, should they get slapped again because they're not prepared or not able to do what I expected 17 years ago? I don't think so. . . .

I think they come with a lot more baggage today than they did 17 years ago—a lot more personal issues, parental guidance issues. . . . We're doing more parenting, and that's just as valuable. (Mr. Ming)

Mr. Liu and Mr. Ming described their assessments as compassionately tailored to the realities of their students' troubled circumstances and consequent skill levels, adaptations not possible with the state's standardized test. Many teachers spoke of efforts to personalize assessment, and many indicated they wanted to implement even more personalized assessment but were prevented by serious limitations on a critical resource—time.

Overall, classroom impact data tended to agree with prior research indicating that testing is having so profound an impact in many classrooms that reform is driving curriculum, a positive effect to the extent that there may now be "less fluff" but negative where pressure to raise test scores eliminates flexibility (Horn, 2002) and focuses on scores rather than on students.

Assessment training for teachers. The teachers' preparedness to meet formal expectations regarding assessment (Washington State Senate, 1992; AFT, NCME & NEA, 1990) appeared to be uneven and inadequate, consistent with wider reports of insufficient teacher training in assessment (Hargreaves, Earl & Schmidt, 2002; Stiggins & Conklin, 1992). Most interviewees, although not all, considered their undergraduate assessment training inappropriate for classroom use. For example, two described their pre-service assessment training as "minimal" (Ms. Vargas, Ms. Walker). Another said she had had only one assessment class in college, which proved unrelated to her content area and which she found to be "useless" for her own teaching (Ms. Grant).

While two teachers reported no assessment training whatever since their initial teacher preparation and many said they had not taken post-graduate assessment courses, others described in-service training in assessment as a "never-ending process" (Ms. Jones) of classes, meetings, seminars, and workshops for local educators and administrators. However, opinions about the quality of professional development in assessment were sometimes no more positive than those regarding college and university assessment courses, one teacher describing assessment training as "a huge inadequacy" (Ms. Nunn). Not only the adequacy but also the appropriateness of the training offered to teachers emerged as suspect. Recent assessment training by one local district, some teachers said, emphasized writing WASL-like questions for implementation in their classrooms "to get [students] used to that type of assessment" (Ms. Park), rather than understanding of measurement principles. In-service training had preempted rather than promoted teacher-developed assessments, reported one teacher who said, "I never developed my own assessment techniques because I was trained by the school and district in the way that they wanted assessment done" (Ms. Apple).

Some teachers identified their colleagues as a more important source of

assessment information that pre-service or in-service training. Two teachers referred to assistance they had received from mentor teachers (Mr. Eggle, Ms. Frank), and one of these derided "new assessment ideas [as] just old ideas draped in new jargon that confuse and threaten older teachers" (Mr. Eggle).

Student impact

The teachers' awareness of the impact of the state test on their students was abundantly evident in their comments. Some indicated that they considered student accountability a commendable state goal. One approving teacher, for example, said, "The WASL is a good thing to hold kids accountable" (Ms. Park). A teacher with a less favorable view of the WASL nevertheless implied that more state testing for the purpose of making promotion and retention decisions was desirable, with a lament that "we do not have any exams to hold kids accountable for moving on to the next level" (Mr. Ming). Another teacher indicated preference for earlier and more frequent imposition of test-based consequences for students in commenting on the inappropriateness of delaying student accountability until high school graduation, saying that he considered it "odd" that the WASL "counts" only for tenth-graders (Mr. Carr). Most teachers, however, expressed serious concerns regarding the WASL's impact on students, as the comments to follow indicate.

Effects of the state test on student self-esteem

Of the teachers interviewed, those who taught at a tested grade or in a tested subject and those who did not both expressed concern regarding the impact of the WASL on student self-esteem. The teachers typically described environments in schools and in classrooms as highly charged during testing windows, one teacher referring to "a lot of stress for both the takers and the administerers of the test" (Ms. Good). Even a teacher who spoke favorably about the test warned that the WASL:

does have too much pressure and overwhelms the students. The scores affect their self-esteem. Nothing is in place [for students who don't meet] the standards. (Ms. Park)

Developmental appropriateness. Data indicated that few teachers considered the state test too easy relative to the content and expectations of their classrooms. One who did said, "[W]hat we expect from students is a lot harder than anything that the WASL tests for" (Ms. Apple). Much more common were concerns that the test was too difficult for some students, to the point of being "not developmentally appropriate for fourth graders. I've seen kids crying about the nightmares they've had over this" (Mr. Felix). A fourth grade teacher expressed the greatest degree of concern about the test, saying:

It is developmentally too difficult for [fourth grade] students. They try so hard when they have no chance of passing. [In the writing portion of the test,] to get a four [the highest score, requires a student to write] better than I could write. This is what they call raising standards. Raising standards means putting it beyond their developmental level and hoping they are going to reach for it. We know that doesn't work. (Ms. Roberts)

The difficulty level of the fourth-grade test and perceptions of its developmental

inappropriateness led some teachers to the conclusion that the test was unfair. Perceptions of inequity were exacerbated for students eligible for special services, for English language learners, and for students with low socioeconomic status.

Effects of the state test on diverse and disadvantaged students

Testing special education students. Some teachers suggested that, for special education students, test time was ill-spent because the WASL offered them "no chance" to demonstrate their knowledge and skills. Even a teacher who approved the test said, "I feel that special ed students should not have to take the test . . . Their time could be better spent on more educational experiences" (Ms. Park). Another teacher objected:

[M]y EMR, which is educable mentally retarded, students have to take the WASL. Learning disabled students have to take the WASL. . . . The EMR students are not going to be successful [on the test], yet we put them through 400 minutes of sweat when they could be having other kinds of experiences. (Ms. Roberts)

Few students could be exempted outright from taking the test, teachers said. However, accommodations were available for students classified as eligible for instructional assistance, but only if the accommodations provided during testing matched ordinary classroom accommodations. Although this general policy sounded reasonable in the abstract, specific restrictions on accommodations rendered it useless, according to one teacher who said:

If a person is learning disabled in writing, and we wanted someone to scribe for them—take dictation—[the student] would have to have that person all year long, every time we had a writing assignment. . . . It can't be just an accommodation for the WASL testing window. Individual students [would] be requiring a lot of time from our [teaching assistants], which we don't have. (Ms. Roberts)

Other accommodations teachers thought might permit documentation of actual achievement were sometimes denied, as described by one teacher recalling a "special needs student [who] could sit down at the computer . . . [where] she had a way of expressing herself" but was not allowed to use the computer when taking the WASL (Ms. Crane).

These teachers' experiences of testing special education students reflected the views of teachers nationally who have objected to state tests as merely providing a new way to show these students they are failures (Horn, 2002). Elsewhere, such perspectives have been brought to bear in legal action, for example, in the 1998 class action lawsuit charging the Indiana state test with unfairness to special education students.

Testing English language learners. The teachers described testing practices for English language learners as no better than those for special education students. A middle school teacher observed:

ESL students only get a one-year exemption from the WASL, which is

not nearly enough [time] to [become] familiar with the language, material, and culture to do well on the test. (Ms Nunn)

A fourth-grade teacher fretted:

They test everyone including kids who have only been [in the U.S.] for a year and a half, so they're taking a test they cannot read. . . . Even though you [might] say, "Oh, they can have assistance," the ESL kids [can only] have the problems read to them verbatim. (Ms. Roberts)

The students would still have to write answers in English.

One teacher who declared that the WASL "doesn't work well when used to assess minorities or special ed students" raised a question of serious practical consequence: "What do we do with the students who cannot pass [the test] year after year and fail to advance?" (Mr. Ochre).

Testing low SES students. Teachers indicated that they considered students in straightened economic and personal circumstances in no less need of consideration than special education students and English language learners. One teacher pointedly predicted, "I'm sure the WASL scores will be best correlated to how much does your mom and dad make economically" (Mr. Ming). In fact, as this teacher intuited, historically, the strongest correlate with scores has been socioeconomic status. The effect of socioeconomic status on test scores was no small matter to the teachers interviewed. For the three-year period 1998-2000, 9-10% of the population in the state of Washington had been considered impoverished (U. S. Census Bureau, 2000). At the time of this study, Vancouver, in the Portland metropolitan area, was suffering from Oregon's highest unemployment rate in the U.S. (Preusch, 2001).

Some teachers poignantly acknowledged increasing levels of economic and social disarray in many families and the consequent calamity in the lives of stricken students. One teacher worried about "how much assistance and guidance do parents provide and are they abusive or intoxicated" (Ms. Grant). No accommodations were available for students suffering the effects of these and other detriments to their real academic opportunities, and no consideration of such background variables were taken into account in calculating their individual achievements as scores on the state test.

Testing students with diverse learning styles. Consistent with the popular theory of multiple intelligences (Gardner, 1983), some teachers noted an inequity derived from discrepancies between the test (content and format) when compared to the different kinds of skills, achievements, and knowledges students might actually possess. Like most standardized achievement testing, the WASL emphasized "logical- mathematical" knowledge and skills over most other types of achievement. Within this theoretical context, these teachers implied that students with strong accomplishments in areas not included on the WASL were unfairly judged non-proficient by a state test that measured a restricted range of achievement.

Teacher impact

Accountability pressures

Most interviewees explicitly recognized that “society wants accountability” (Ms. Good) and that “raising the standard would raise the credibility of the American public school system” (Mr. Carr). They were keenly aware of public scrutiny of WASL scores published in local newspapers.

Almost unanimously, more pervasively than has been reported nationally (Abrams, 2002), the teachers, even those who described themselves as relatively unaffected by testing and test pressures, noted societal pressures related to scores and accountability. Fewer than one-fourth of the teachers interviewed, most of these in untested grades, indicated that the WASL had little impact on them. One said that his plan for avoiding test-related demands was to retire so that he would be “long gone” before it was necessary for him to align his curriculum with the state test (Mr. Liu).

No teacher in this study objected to accountability *per se*, but several teachers expressed frustration at being held accountable for test results when student performance depended not only on teaching but also on factors beyond teacher control, factors they listed as including class size, student ability, primary language, eligibility for special services, socioeconomic status, transience, family difficulties, and motivation. “Your teaching [will] eventually be judged by the kids who blow it off,” fumed one teacher (Ms. Good). Another recognized teacher vulnerability where “students perform badly on the WASL intentionally to make a point” of their own objections to the test (Mr. Dustin).

While several teachers considered the impact of state testing meritorious, most expressed concern regarding the appropriateness of the state's prioritization of test scores in reckoning school accountability. This was consistent with other research findings that teachers do not oppose standards or accountability, but most disagree with current uses of test scores for school accreditation (Abrams, 2002; Shore, 2002).

Effects on classroom instruction and assessment

Some teachers in this study described classroom effects similar to reported trends indicating that state tests “deform curricula” (Schoenfeld, 2002). The teachers identified such things as how to fill in the bubbles on answer sheets and how to follow prompts as examples of local WASL preparation activities which took time away from regular teaching and learning.

The extent of curriculum displacement alarmed some teachers, one of whom said, “When we do the WASL, our school is in chaos for the entire time. I lose a month of teaching. It affects the whole school” (Ms. Doe). Others reported that “test prep” consumed as much as five or six months in a tested grade. One teacher complained that the test overwhelmed classroom instruction even in untested grades:

I guess I'm one of many teachers who feel there is so much emphasis on the WASL that it has almost become the focus of our teaching. I'm

not real comfortable with that. . . . [F]or the fourth graders, the minute they enter fourth grade, they're hearing about the WASL and how they have to do well on the WASL. . . . But even at third grade, I find myself saying to students, "This is the type of question that you will have on the WASL when you are in fourth grade." . . . [W]e're just so test-oriented that we've kind of lost sight of what education truly is. (Ms. Quinn)

Some teachers expressed resistance to reallocating instructional focus and time for test preparation, one saying, "I can't just prepare my students to take the WASL. It's not the only thing that should be assessed" (Ms. Brush). Another, who complained that the WASL, a standardized test, "doesn't measure anything that we teach our kids," declared, "we are not willing to change because what we do for our kids is what they need" (Ms. Apple).

But many fell into line, some with misgivings or under duress. A high school science teacher, for example, had reluctantly added earth science to her curriculum because it was found in the EALRs, although it was outside her specialty (Ms. Walker). Another teacher reported a shift away from thematic instruction and toward a fragmented approach to the curriculum as she "hit subject matter individually while constantly checking and re-checking the EALRs" (Ms. Jones). One reported her teaching was "becoming more canned" (Ms. Hallo). Another described herself as physically displaced in her own classroom, to some extent, by tutors brought in to ensure her students were prepared for the WASL (Ms. Doe).

The amount of class time devoted to external assessment was not limited to preparation for and administration of the WASL. Teachers reported at least eight additional standardized achievement tests, seven developed and marketed by big-name commercial testing corporations, in use in their schools or districts.

Not only instructional practices, but classroom assessment practices, too, were increasingly pressed into the WASL mold, data indicated. For example, one teacher said she had reorganized her students' portfolios "to match the EALRs." As a member of her district's "assessment training team," led by an official from the state education agency, she was "learning how to write a practice test similar in format to the WASL" as her district developed "a WASL-like practice tests for second graders" (Ms. Park).

Teacher perspectives

Slightly less than half of the teachers interviewed expressed approval of the WASL or of some aspects of it. Two praised the test's emphasis on "process," one adding that this emphasis was "good because we're trying to make a more fair assessment" (Ms. Hand), and the other praising partial credit given to students who showed workable math procedures even when answers were ultimately incorrect (Ms. Doe). The latter also approved the WASL's authentic eliciting of "the same skills [students] use in real life" (Ms. Doe).

Most expressions of approval included qualifications. For example, one teacher said the WASL was "probably a good thing" (Mr. Inder), another that she felt positive "for the most part" (Ms. Frank), and another that it was a good thing to hold students accountable although "improvements to the test" were needed (Mr. Alder).

One said the WASL "can assess some [students]. I think that it cannot assess all" (Ms. Roberts). Content limitations were noted by a teacher who observed that the test content was "not the only thing that should be assessed" (Ms. Brush). One teacher who approved the test distinguished between its quality and its utility: "I like the test. I just don't know how it should be used" (Mr. Carr).

The most positive opinions were offered by two teachers involved in developing either a practice WASL-like test or a rubric to standardize the assessment of student writing. One of the two had been a member of a district assessment team for four years, "so long it has become a part of me, and I have begun to buy into it." Even so, her praise of the WASL was qualified: "The test is still new. The kinks have to be worked out" (Ms. Park). The other, who said she had helped develop Six Trait Writing Assessment, indicated that she had changed her teaching in response to the state standards, which she considered congruent with her beliefs and practice, but that she disapproved of the WASL as "much too narrow a device" (Ms. Underwood). This small (n=2) positive correlation between individuals' involvement with test development and their approval of the WASL was consistent with findings from a nation- wide study:

[T]he promotion of greater receptivity towards change at a local level, which might entail teacher knowledge about the reform, shaping attitudes toward reform objectives, or providing greater "how- to" knowledge instrumental for implementing change. . . . appears to be a likely mechanism through which this policy reform operates. (Swanson & Stevenson, 2002, p. 15)

However, it was unclear whether local data suggested that teachers' close scrutiny of the testing system led them to appreciate the WASL or, alternatively, suggested that involving teachers in development of standards and tests habituated and coopted them.

Most teachers took issue with the test, the most virulent wording coming from one who called the WASL "stupid" (Mr. Banks) and another saying, "I despise it" because of its counterproductiveness regarding learning and its deprofessionalization of teachers (Mr. Twain). Complaints centered on negative impacts to curriculum, students, classrooms, and schools, as previously detailed, articulating questions and concerns about equity and developmental appropriateness, as noted earlier, and about validity, scoring, expense, and volatile state policies and requirements, to be discussed in the next section.

Teachers' objections also included lack of useful feedback in the reporting of test results. "It would be nice for kids to get the tests back and see the mistakes that were made so that they could focus on their weaknesses," said one of two teachers (Ms. Vargas) who objected to delayed notification of WASL results. The other estimated the delay as "six months" after test administration, too late for corrective instruction. When results did arrive, she complained, there were further obstructions to fulfillment of the state goal that testing help identify remediation needs:

As far as I can tell, there has been no interpretation of what failing test results mean. . . . [And] I am not allowed to keep the test results. I am

only allowed to see them for a short time because they are locked up. I don't know if that's in every school or just in this school. (Ms. Roberts)

One teacher objected to the content of a specific item, saying he had "lost respect" for the WASL after publicity about a tasteless question that referred obliquely to a notorious trial involving a teacher's alleged seduction of a student (Mr. Ochre).

Overall, local teachers' concerns about state testing closely matched those of their colleagues nationally: fairness, timeliness of feedback, diagnostic value of test result reports, single-shot testing, pacing in classrooms, the number of tests, extraneous factors that affect scores, and pressure to cover all the standards (Shore, 2002).

Validity concerns

Validity through multiple measures. Although no teacher used technical terms in responding to interview questions, analysis of data from the perspective of traditional psychometrics revealed strong practitioner understanding of important measurement concepts and principles, particularly regarding validity. Teachers made clear their intuitive understanding of the injunction to use multiple measures in order to make valid inferences and decisions regarding a student's achievement, as specified in the *Standards for Educational and Psychological Testing*:

Standard 11.20. In educational, clinical, and counseling settings, *a test taker's score should not be interpreted in isolation*; collateral information that may lead to alternative explanations for the examinee's test performance should be considered. (AERA, AEA & NCME, 1999, p. 117, emphasis added).

Similarly, the standards for educational accountability systems developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Consortium for Policy Research in Education (CPRE) prominently and succinctly state, "Decisions about individual students should not be made on the basis of a single test" (Baker, Linn, Herman & Koretz, 2002, p. 3). The American Evaluation Association, in its first public policy pronouncement, has counseled against "simplistic application of single tests or test batteries to make high stakes decisions about individuals and groups [which] impede rather than improve student learning" (2002, unpaginated). The National Association for the Education of Young Children (NAEYC) has issued a position statement declaring:

Decisions that have a major impact on children such as enrollment, retention, or assignment to remedial or special classes should be based on multiple sources or information and should *never* be based on a single test score. (NAEYC, 1988, emphasis added)

The teachers interviewed spoke of their own multiple measures as providing more accurate portrayals of their students abilities than the state test could provide. The WASL, said one, was merely "one window into a child for one week. As a teacher, I can tell you about their growth as a student" (Ms. Hand). All the teachers agreed that frequent and varied methods were needed to understand and represent accurately the diverse accomplishments of their students.

Washington state relied essentially on the WASL, (Note 6) although the awareness of the importance of multiple measures was indicated in such public statements as the following:

No single test can tell you everything about a child's performance. Looking at information from a variety of tests and assessment tools remains the best way for parents and classroom teachers to really see how well individual students are learning. (Office of the Superintendent of Public Instruction website, www.k12.wa.us, June 20, 2002)

Construct validity. Some of the teachers interviewed explicitly challenged the WASL's construct validity, questioning whether the test did, in fact, test what it purported to test—the *construct of student achievement*. For example, one teacher said, "There is too much confusion about what it is actually trying to measure" (Ms. Nunn). When scores reflect things besides the intended construct (i.e., *rival constructs*), test results can be misleading, either exaggerating achievement or denying due credit.

The math problem-solving section of the WASL was perceived by some interviewees as troublesome on these grounds, requiring students to explain their solution procedures. Teachers reported that many students who were good at math but weak in writing were unfairly penalized. Said one teacher, "Even if they can explain their thinking and they have the answer right, they get marked down because of their writing skills" (Ms. Hallo). Problems related to rival constructs were not limited to writing requirements in the math test. Teachers suggested several rival constructs actually being measured rather than (or in addition to) the intended construct, student achievement, in saying:

Rival construct—socioeconomic status of individual students: "I'm sure the WASL scores will be best correlated to how much does your mom and dad make economically. . . . Socioeconomic status is the greatest predictor of student success." (Mr. Ming)

Rival construct—personal difficulties: "There are other variables that go into testing—a baby, a job, living in their cars. These affect test performance." (Ms. Hand)

Rival construct—intelligence: "I firmly believe that WASL performance is not only affected by teaching but also [by] cognitive abilities which, to a certain extent, are innate." (Mr. Exeter)

WASL scores are used not only as measures of the achievements of individual students but "to evaluate instructional practices" and "to hold schools accountable for student learning" (Washington State Senate, 1992, p. 10). For this reason, the validity of inferences on *the construct of school or educational quality* emerged as relevant in the analysis. Several teachers' comments indicated realization that a school's test results might indicate not the quality of its educational program delivery but, rather, the characteristics of its student body, including student motivation and especially affluence:

Rival construct—student motivation: "[Some] students perform badly on

the WASL intentionally to make a point." (Mr. Dustin)

Rival construct—socioeconomic status of school population: "[A school in my district] traditionally has been at the top but, since we've redistricted, they had a huge influx of students from the lower echelon housing and economic development. That has changed their dynamics. They didn't do as well as they had hoped [on WASL scores]. . . . [S]tudents who are socioeconomically deprived don't do as well." (Ms. Roberts)

Content validity. In describing the test as "much too narrow a device" (Ms. Underwood), one teacher implied that not only construct validity but also content validity was at issue, that the content of the test did not sufficiently represent the content of the intended domain (e.g., the English-language arts test did not fully represent the domain of English-language arts).

Instructional validity. Relatedly, some teachers indicated that instructional validity—the match between what is taught and what is tested—was faulty, one observing that the WASL "doesn't measure anything that we teach our kids" (Ms. Apple). Another teacher complained that test content was insufficiently aligned with the curriculum:

I like the fact that people are accountable for teaching certain curriculum, but the assessment part needs work. There is a lot of mismatch between the curriculum and what the WASL is testing. (Ms. Walker)

Scoring concerns

Concerns about the scoring of the state test were also raised. Interpreting a student's written explanation requires professional skill, experience, knowledge of child development, and sometimes knowledge of the particular child, according to one teacher who said:

The WASL is graded by people with no idea of knowing what good communication is for that child. There's a greater possibility for a disconnect that's unfair for the student. (Ms. Underwood)

The importance of accurate interpretation of text generated by children was not limited to tests of reading and English-language arts. As noted earlier, there were also concerns that students' math achievements might not be fully credited because of the scoring of verbal explanations: "One could be good at math but can't explain their thinking. They would be judged as not passing the test" (Ms. Park).

Two teachers complained that some schools were inappropriately penalized because of regulations related to student scores of zero. One reported that the state had required GED students be classified as sophomores and prohibited them from taking the WASL, then had counted the "lack of scores" from these students against her school, the county's GED school, artificially lowering the school's results (Ms. Apple).

Test expense

A few teachers expressed concern regarding the cost of testing, one preferring a "standardized test which is cheaper and faster" (Ms. Roberts) than Washington's current standards-based (and standardized) test with its performance assessment sections. Another hoped "the state isn't wasting millions of dollars" (Mr. Alder).

Changing state policies and requirements

Some teachers approved the WASL and expected that testing would always be part of the educational system, but one worried about the diversion of resources to the WASL if it proved merely to be "some fad that won't be around long" (Mr. Alder). Another expected no more:

The WASL is just another one of those things that's going to come, and it, too, shall pass. I haven't changed what I teach or how I teach because, as a conscientious professional, I've looked at what students should know in terms of biology. (Mr. Ming)

In fact, changes to state accountability and testing policy have been enacted "almost every year" (OSPI website, www.k12.wa.us, June 20, 2002) since SSB 5953 in 1992 (see Table 4). Frequent changes, creating layers of increasing and sometimes conflicting requirements, can be seen across the country as state testing programs have increased during the last decade, partly in response to federal requirements regarding Title 1 funding, and in the new federal requirement to test all children in grades 3-8 every year. "Policy hysteria" (Stronach & Maclure, 1996) is a term which has been given to frequent, overlapping policy changes in general (i.e., not necessarily related to testing).

Table 4
Summary of state statutes regarding Washington's education reform initiative

year	bill	effect
1992	SSB 5953	established the framework for education reform and the Commission on Student Learning (expired 1999), providing for the development of the Essential Academic Learning Requirements (EALRs) and a new assessment system.
1993	ESHB 1209	resulting from work by the Governor's Council on Education Reform and Funding (GCERF), established new learning goals and Student Learning Improvement Grants (SLIG) and other programs to help educators help students meet new standards.
1994	ESHB 2850	established requirements pertaining to character traits and values.
1995	SSB 5169	made relatively minor changes to prior law.

1997	ESB 6072	established a timeline for assessment development.
1997	ESHB 2042	established a grade 2 reading assessment.
1998	ESHB 2849	required district school boards to establish reading improvement goals. Also, a grade 4 NRT was moved to grade 3. Also, the legislature provided funds for professional development, instructional materials, and schools with reading programs involving volunteer mentors.
1999	ESHB 5825	made changes to the NRTs and modified the assessment implementation timeline.
1999	SSB 5418	established the Academic Achievement and Accountability Commission, established mathematics goals, and created several new assistance programs.
2002	ESB 6456	authorized the A+ Commission to set performance improvement goals for all students (e.g., economically-disadvantaged students, limited English proficient students, students with disabilities, and students from disproportionately underachieving racial and ethnic backgrounds) and to establish high school graduation rate goals and dropout reduction goals for grades seven through twelve.

Source: Website of the Office of the Superintendent of Public Instruction, State of Washington, www.k12.wa.us, June 20, 2002

Findings

Variations among the perspectives of the 31 teachers interviewed signal continuation of a robust collective struggle to understand and improve education. The variations also evidence the kind of diversity and local control which many have considered traditional strengths of American schooling. The contrasts were so dramatic that two interviewers were "not shocked but stymied" in trying to analyze the range of opinion expressed by the four teachers they had interviewed—perceptions of the state test ranging from approval to ignorance to objection, perceptions expressed with a range of emotions from candor to arrogance to wariness, perceptions varying as to whether the teachers' own assessment practices should follow state mandates or personal beliefs.

Several interviewers expressed surprise that teachers were not more negative about state testing but, instead, that some had offered positive comments or described the state test as a tool to help their teaching. Other interviewers were taken aback by teachers' deep distress about the test and its implications, two interviewers writing, "We feel as overwhelmed as the teachers." Overall, the teachers in this study, like teachers across the country (Shore, 2002), appeared to be adapting and trying to make things work. From the data they provided, four main findings emerged.

(1) *The teachers did not fear accountability but opposed accountability based on a*

single-shot test. Their opposition reflected better understanding of the important principle of multiple measures than was manifested in the state accountability policy. Teachers' intuitive, experiential understanding—sometimes referred to as "practical wisdom"—appeared to be stronger in this regard than the formal understanding of state officials and their testing contractors and consultants who had implemented a test-driven accountability system with heavy reliance on the WASL.

(2) *The WASL was not appropriate for children who were eligible for special services, who were non-proficient speakers in English, or who were living in impoverished or marginal situations*, according to teachers who worked with them day-to-day. Teachers indicated that *the state test ensured that these children would not only be left behind but also pressured and punished for factors beyond their control. Individual student scores aggregated and reported as school scores similarly pressured and punished teachers for factors beyond their control*, said some.

(3) Teachers repeatedly claimed *classroom assessments were more informative but sidelined by the state tests*. One teacher, for example, referred to the WASL as "one measurement done during a short period of time that provides a little glimpse of the students, [whereas] I have them all year so I have a better perspective on them" (Ms. Hand). Teachers already understood the message researchers have been trying to share with policymakers, for example:

Once-per-year accountability tests can't do the job of day-to-day, week-to-week pupil diagnosis. . . . What large-scale assessment can't do is document in sufficient detail the what and how of student understandings. (Shepard, 2002)

Policy-makers need to support the development of new assessments and to avoid reliance on single tests. They should shift resources from large-scale assessment to classroom assessment. (Pellegrino, 2002)

(4) *While some teachers appreciated the focus provided by state standards and testing, other teachers were troubled by the test's replacement of teachers' professional judgment*:

The WASL goes against everything we know about learning and takes assessment out of the hands of educators and puts it into the hands of a corporate organization out for profit. (Mr. Twain)

The WASL is robbing me of my professional judgment and replacing learning with inappropriate practices. (Ms. Quinn)

If "inappropriate practices" are the result of state testing, teachers should resist. Although some have blamed teachers' insufficient resistance for the current wave of high-stakes testing (Popham, 2001), some teachers in this study indicated staunch resistance to Washington's state test within their classrooms, their clearest spheres of influence and the location of their primary responsibilities.

High stakes testing represents a mechanism to ensure local compliance to policy initiatives typically described as "reform." Efforts to comply were evident in this study. It nevertheless seems unlikely that centralized, top-down, state control can

lead to better education, as implied by the term "reform" (see Fullan, 1991; Sarason, 1990) when it simultaneously deprofessionalizes teachers by usurping their authority and opportunity to plan and implement educational opportunities for their students. In a postmodern era skeptical of grand plans and centralized management, it is worth considering whether forcibly turning teachers into technicians, a return to the previous century's "technological perspective" for controlling education (Hargreaves, Earl & Schmidt, 2002) or "technicist approach" for making education efficient (Gillman, 2002), is more likely to re-form education in a detrimental rather than in an improved manner.

Conclusion

During the data analysis phase of this study, Washington state superintendent Terry Bergeson publicly and plaintively remarked that, as a former school counselor she was not initially an advocate of large-scale testing, but "we need data" (2001). Four months later, a district administrator from Kansas City complained, "We're drowning in data but parched for information—and the questions are cosmic" (Wright, 2002). This study suggests that teachers, who face cosmic questions in the microcosms of their classrooms, are a source of information that policy-makers would be well-advised to heed.

It is no small matter that more than two-thirds of U.S. teachers consider their state tests not worth the investment (Abrams, 2002) and that some teachers are leaving the profession because of test pressure (Gillman, 2002). Teachers are crucial to educational reform not only for the well-known reason that top-down mandates succeed only with bottom-up buy-in from implementers (Fullan, 1991; Sarason, 1990). In addition, teacher perspectives are key to implementing *reasonable* accountability (Shore, 2002) because it is teachers who bring together understanding of children, their achievements, and how to assess them. Teachers' understanding of assessment, despite deprivation of strong formal training, has been too long underestimated. The data clarified a critical difference of emphasis: teachers' focus on "testing *students*" and state or external focus on "*testing* students."

Moreover, understanding teachers' experiences and perspectives helps to explain research findings regarding "perverse incentives" related to state tests, such as teachers' unwillingness to accept or keep positions in low-scoring schools that most need their expertise and energies (Trent, 2002; see also Lankford, Loeb & Wyckoff, 2002). At a time when teacher shortages and high turn-over rates are a matter of concern in Washington state, careful consideration is needed in policy-making circles regarding the impact of test-driven accountability on teacher recruitment, retention, and job satisfaction.

Notes

1. Only 19 states ever reached full compliance (*Education Week*, April 17, 2002, p. 29), and the national system of tests was never developed.
2. Prior to NCLB, NAEP was voluntary for states.
3. The date for making passing the WASL a graduation requirement has been extended to 2008.

4. For permission to use their interview data and for review of a draft of this manuscript, the authors wish to thank Kevin Crouch, Candace Dawson, Patrick Dowell, Daniel Getty, Jeff Herzog, Stephen Klauer, Karissa Lowe, Jennifer Megli, Mark Muckerheide, Mary Nelson, Wayne Storer, Debra Tidd, and Chad Towe. The authors also thank Marv Alkin of UCLA for review and comments regarding a draft of the article.
5. Since each graduate student interviewed two teachers, there should have been an even number of teachers in the sample. However, by chance and without realizing it, two students chose and interviewed the same teacher.
6. The state also mandated administration of the Iowa Test of Basic Skills (ITBS) in grade 3 reading and math and in sixth grade reading, language arts, and math; and of the Iowa Test of Educational Development (ITED) in grade 9 reading, language arts, math, and an interest inventory. (Source: www.k12.wa.us)

References

- Abrams, L. (2002, April). Multi-state analysis of the effects of state-mandated testing programs on teaching and learning: Results of the national survey of teachers. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- American Evaluation Association Task Force on High Stakes Testing. (2002). Position statement on high stakes testing in preK-12 education. Fairhaven, MA: AEA.
- American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Washington, D.C.: Authors.
- Baker, E. L. (2002, April). Validity issues for accountability systems. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). Standards for educational accountability systems. *CRESST Line*, Winter, 1-4.
- Bergeson, T. (2001, December 6). Washington reform update and implications for the future. Presentation to the Washington State Assessment Conference, Seattle, WA.
- Bond, L. A., Braskamp, D., van der Ploeg, A., & Roeber, E. (1996). *State student assessment programs database, school year 1994-95*. Oak Brook, IL: Council of Chief State School Officers and the North Central Regional Educational Laboratory.
- Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton-Mifflin.
- 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.
- Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Denzin, N. K. (1997). *Interpretive ethnography: Ethnographic practices for the 21st century*. Thousand Oaks, CA: Sage.
- Denzin, N. K. & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Education Week* (2002, April 17). 1994 ESEA: The state of state compliance. Authors, p.. 29.
- Elmore, R. (2002, April). Stakes for whom? Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

- 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.
- No child left behind (NCLB), reauthorization of the Elementary and Secondary Education Act, Public Law 107-110 (2001).
- 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.
- Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gewertz, C. (2002, April 10). Low-scoring charter school to shut down in Chicago. *Education Week*, p. 4.
- Gillman, C. (2002, April 26). From blooming flowers to marching soldiers: A case study of one kindergarten teacher. Colloquium, Washington State University Vancouver.
- GOALS 2000: Educate America Act, Public Law 103-227 (1994).
- 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.
- 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.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8 (41). (<http://epaa.asu.edu/epaa/v8n41/>)
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39 (1), 69-95.
- Hodder, I. (1994). The interpretation of documents and material culture. In Denzin, N. K. & Lincoln, Y. S. (Eds.), *Handbook of qualitative research* (pp. 403-412). Thousand Oaks, CA: Sage.
- Horn, C. (2002, April). Multi-state analysis of the effects of state-mandated testing programs on teaching and learning: Results of the fieldwork studies. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24 (1), 37-62.
- LeCompte, M. D. & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego: Academic Press.
- Lewis, S. (2002, April). What will be the effects on assessment and accountability in local school districts of the "no child left behind" legislation? Presentation to the annual meeting of the National Council of Measurement in Education, New Orleans, LA.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29 (2), 4-16.
- Mabry, L. (2002). In living color: Qualitative methods in educational evaluation. In D. Nevo & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*. Boston: Kluwer-Nijhoff.
- Mabry, L. (1999). *Portfolios plus: A critical guide to alternative assessments and portfolios*. Thousand Oaks, CA: Corwin Press.
- Mabry, L. (1998). Case study methods. In H. J. Walberg & A. J. Reynolds (Eds.), *Evaluation research for educational productivity* (pp. 155-170). Greenwich, CT: JAI Press.
- Mabry, L., Aldarondo, J., & Daytner, K. (1999). Local administration of state-mandated performance assessments: Implications for validity. Paper presentation to the annual meeting of the American Educational Research Association, Montreal, Canada.
- Mabry, L. & Daytner, K. G. (March, 1997). State- mandated performance assessment. Paper

- presentation to the annual meeting of the American Educational Research Association, Chicago, IL.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62 (3), 279-300.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Meyer, L., Orlofsky, G. F., Skinner, R. A., & Spicer, S. (2002). The state of the states. In *Quality counts 2002: Building blocks for success* (a report on education in the 50 states by the Editorial Projects in Education). *Education Week*, 21 (17), 68-92.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age, adopted November 1987. *Young Children*, 43 (3), 42-47.
- Neuman, S. (2002, April). The Bush accountability and assessment agenda: New opportunities and challenges. Presentation to the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Orlofsky, G. F. & Olson, L. (2001). The state of the states. In *Quality counts 2001: A better balance* (a report on education in the 50 states by the 2001 Editorial Projects in Education). *Education Week*, 20 (17), 86-92, 94-100, 102- 106.
- Oregonian*. (2002, November 29). Nearly all Washington schools fail to hit goals. Portland, OR: Authors, p. B2.
- Pellegrino, J. (2002, April). Assessment and learning: Issues highlighted in the NRC report "Knowing what students know." Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Pomplun, M. & Capps, L. (1999). Gender differences for constructed-response mathematics items. *Educational and Psychological Measurement*, 59 (4), 597-614.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Preusch, M. (2001, December 15). National briefing Northwest: Oregon unemployment rate rises. *New York Times*, p. 14
- Rubin, H. J. & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.
- Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it's too late?* San Francisco: Jossey-Bass.
- Schoenfeld, A. (2002, April). "This is just a test!" Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- 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.
- Shepard, L. A. & Smith, M. L. (1988). Escalating academic demand in kindergarten: Counterproductive policies. *Elementary School Journal*, 89 (2), 135-145.
- Shore, A. (2002, April). Optimizing the validity and value in the public debate over testing as a tool in educational reform. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20 (5), 8-11.

- Smith, M. L. & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10 (4), 7-11.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7 (2): 5-8.
- 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.
- Stiggins, R. J. & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: SUNY Press.
- Stronach, I. & Maclure, M. (1996). Mobilizing meaning, demobilizing critique? Dilemmas in the deconstruction of educational discourse. In *Cultural Studies* (vol. 1, pp. 259- 276). Greenwich, CT: JAI Press.
- Swanson, C. B. & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, 24 (1), 1-27.
- Trent, W. (2002, April). The policy implications of federally mandated annual testing. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- U. S. Census Bureau. (2001). Percent of people in poverty by state. Online at <http://www.census.gov/prod/2001pubs/p60-214.pdf>.
- 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.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.
- Wright, D. D. (2002, April). Who did we miss, and why? Factors associated with non-participation of general education students in standardized assessments. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Yakimowski, M. (2002, April). What will be the effects on assessment and accountability in local school districts of the "no child left behind" legislation? Presentation to the annual meeting of the National Council of Measurement in Education, New Orleans, LA.

About the Authors

Linda Mabry is an associate professor at Washington State University Vancouver, where she specializes in assessment of student achievement, program evaluation, and qualitative research methodology, and a member of the boards of the American Evaluation Association and the National Center for the Improvement of Educational Assessment.

Jayne Poole is a graduate student in Education at Washington State University Vancouver, where she is researching the reading- writing connection, and a kindergarten teacher of eleven years in Longview, Washington.

Linda Redmond, a recent Masters in Education graduate of Washington State University Vancouver, has taught in the public schools of Washington state for twenty-two years and is currently an elementary music specialist in Longview, Washington.

Angelia Schultz, a newly certificated teacher with a BA in English and a graduate student in Education at Washington State University Vancouver, currently works as a substitute teacher. Her interests include the consequences of high stakes testing and test- driven accountability.

The World Wide Web address for the *Education Policy Analysis Archives* is
epaa.asu.edu

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, [Gene V Glass, glass@asu.edu](mailto:glass@asu.edu) or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

[Michael W. Apple](#)
University of Wisconsin

[Greg Camilli](#)
Rutgers University

[Sherman Dorn](#)
University of South Florida

[Gustavo E. Fischman](#)
California State University–Los Angeles

[Thomas F. Green](#)
Syracuse University

[Craig B. Howley](#)
Appalachia Educational Laboratory

[Patricia Fey Jarvis](#)
Seattle, Washington

[Benjamin Levin](#)
University of Manitoba

[Les McLean](#)
University of Toronto

[Michele Moses](#)
Arizona State University

[Anthony G. Rud Jr.](#)
Purdue University

[David C. Berliner](#)
Arizona State University

[Linda Darling-Hammond](#)
Stanford University

[Mark E. Fetler](#)
California Commission on Teacher Credentialing

[Richard Garlikov](#)
Birmingham, Alabama

[Aimee Howley](#)
Ohio University

[William Hunter](#)
University of Ontario Institute of Technology

[Daniel Kallós](#)
Umeå University

[Thomas Mauhs-Pugh](#)
Green Mountain College

[Heinrich Mintrop](#)
University of California, Los Angeles

[Gary Orfield](#)
Harvard University

[Jay Paredes Scribner](#)
University of Missouri

[Michael Scriven](#)
University of Auckland

[Robert E. Stake](#)
University of Illinois—UC

[Terrence G. Wiley](#)
Arizona State University

[Lorrie A. Shepard](#)
University of Colorado, Boulder

[Kevin Welner](#)
University of Colorado, Boulder

[John Willinsky](#)
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
[Roberto Rodríguez Gómez](#)
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

[Adrián Acosta \(México\)](#)
Universidad de Guadalajara
adrianacosta@compuserve.com

[Teresa Bracho \(México\)](#)
Centro de Investigación y Docencia
Económica-CIDE
bracho dis1.cide.mx

[Ursula Casanova \(U.S.A.\)](#)
Arizona State University
casanova@asu.edu

[Erwin Epstein \(U.S.A.\)](#)
Loyola University of Chicago
Eepstein@luc.edu

[Rollin Kent \(México\)](#)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de
México
javiermr@servidor.unam.mx

[Humberto Muñoz García \(México\)](#)
Universidad Nacional Autónoma de
México
humberto@servidor.unam.mx

[Daniel Schugurensky \(Argentina-Canadá\)](#)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

[Jurjo Torres Santomé \(Spain\)](#)
Universidad de A Coruña
jurjo@udc.es

[J. Félix Angulo Rasco \(Spain\)](#)
Universidad de Cádiz
felix.angulo@uca.es

[Alejandro Canales \(México\)](#)
Universidad Nacional Autónoma de
México
canalesa@servidor.unam.mx

[José Contreras Domingo](#)
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

[Josué González \(U.S.A.\)](#)
Arizona State University
josue@asu.edu

[María Beatriz Luce \(Brazil\)](#)
Universidad Federal de Rio Grande do
Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

[Angel Ignacio Pérez Gómez \(Spain\)](#)
Universidad de Málaga
aiperez@uma.es

[Simon Schwartzman \(Brazil\)](#)
American Institutes for
Research—Brazil (AIRBrasil)
simon@sman.com.br

[Carlos Alberto Torres \(U.S.A.\)](#)
University of California, Los Angeles
torres@gseis UCLA.edu

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University