

**Combining Surveys and Case Studies
to Examine Standards-Based Educational Reform**

CSE Technical Report 565

Brian Stecher
CRESST/RAND

Hilda Borko
CRESST/University of Colorado at Boulder

May 2002

Center for the Study of Evaluation (CSE)
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Sciences
University of California, Los Angeles
301 GSE&IS, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Project 1.5 The Effects of Standards-Based Assessments on Schools and Classrooms

Brian Stecher, CRESST/RAND, and Hilda Borko, CRESST/University of Colorado at Boulder, Project Directors

Copyright © 2002 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

ACKNOWLEDGMENTS

This research is part of a larger project, *The Effects of Standards-Based Assessments on Schools and Classrooms*, conducted through CRESST.

We would like to thank other members of the research team—Shelby Wolf, Genet Simone, and Kay Uchiyama (CU Boulder); and Tammi Chun, Karen Ross, and Sandy Chien (RAND)—who participated in design, data collection, and analysis for the larger project. We thank our colleagues, Kathryn Borman, Larry Suter, and James Spillane, for their comments on the paper. We extend special thanks to the principals and teachers at the schools that are featured in this paper. Without their support, this project would not have been possible. An earlier version of this paper was presented at the annual meeting of the American Educational Research Association, Seattle, Washington, April 2001.

**COMBINING SURVEYS AND CASE STUDIES
TO EXAMINE STANDARDS-BASED EDUCATIONAL REFORM**

**Brian Stecher
CRESST/RAND**

**Hilda Borko
CRESST/University of Colorado at Boulder**

Abstract

It is becoming more common to use multiple research methods when studying large-scale school reforms. For example, over the past 5 years the authors combined statewide teacher surveys and school case studies to examine the impact of standards-based educational reforms in Kentucky and Washington. This paper uses examples from the study of the effects of the Washington education reform to explore how these methods can be used in complementary ways. It presents specific examples of the benefits of using both methods and makes recommendations for more effective integration of case study and survey methods in the future.

The question that motivates this paper is simple: How can surveys and case studies be used in combination in the study of school reform to maximize the usefulness of the data they provide? The answer is more complex. Our experience studying standards-based school reform in Kentucky and Washington suggests that there is potentially great power in combining these two approaches, but we have yet to fully realize that power.

We came to this question as much out of personal interest as through the thoughtful study of methodological options. Speaking metaphorically, we found ourselves in the same place at the same time, bearing different research tools. To be more concrete, we were both interested in how standards-based state reforms with high-stakes testing components affected schools and classrooms. Hilda Borko's interest grew out of many years of studying teacher learning (Borko, Davinroy, Bliem, & Cumbo, 2000; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Putnam & Borko, 2000). Brian Stecher's interest grew out of several large-scale investigations of assessment and its impact on practice (Koretz, Stecher, Klein, & McCaffrey, 1994; Stecher, 1998). Although we work at different institutions in different time zones,

under the auspices of the national Center for Research on Evaluation, Standards, and Student Testing (CRESST), we had the opportunity to combine our interests in a coordinated investigation of the impact of standards-based reform on schools and classrooms. The use of the word *coordinated* rather than *integrated* is the subject of this paper.

Background

For the past five years we have been studying the effects of state-mandated, standards-based assessments on schools and classrooms in Kentucky and Washington. In general, such statewide reforms rely upon principals and teachers to translate the goals embodied in the standards and assessments into practice. Yet, states generally give only limited attention to the processes through which schools foster assessment-based change and teachers adapt the classroom instructional environment to accommodate new curriculum standards and broadened achievement expectations. Although they are generally overlooked in policy making, these processes are essential for effective implementation. As Fullan and Miles (1992) noted, “local implementation by everyday teachers, principals, parents, and students is the only way that change happens” (p. 752).

To address this gap in understanding, our research focused on two broad questions: What are the effects of recent statewide assessment reform on school structure and organization, classroom practices, and student outcomes? What combination of factors explains the differential patterns of success within and across schools and states? We focused our investigations on factors that might explain successful schools and classrooms, such as incentives for change, staff development efforts, local support networks, teachers’ and students’ perceptions of the assessment, and the ways in which principals and teachers responded to information from various sources.

Methodological Issues

One of the features that distinguished our research from past studies of testing reforms was its use of survey and case study Methods to provide both breadth and depth of analysis within a single investigation. We hoped to benefit from the advantages of the two methods highlighted by Patton (1990):

The advantage of a quantitative approach is that it is possible to measure the reactions of a great many people to a limited set of questions, thus facilitating comparison and statistical aggregation of the data. This gives a broad, generalizable set of findings presented succinctly and parsimoniously. By contrast, qualitative methods typically produce a wealth of detailed information about a much smaller number of people and cases. This increases understanding of the cases and situations studied but reduces generalizability. (p. 14)

In this spirit, we developed surveys to identify broad patterns of impact within each state and to explore differences in practice among teachers and schools. In investigations of standards-based reform efforts, surveys are good at examining implementation events, topic coverage, and relatively common aspects of instruction, but they are less effective at capturing many of the important instructional features of these reforms. We complemented the surveys with case studies that explored, in a small number of schools and classrooms, the complexity of factors that interact to determine the differential success of the reform efforts. Case studies can reveal much more about teachers' understanding and the evolution of instructional practices, although they are far too labor intensive to use on a large scale.

Other researchers have recognized the important role that case studies play in developing "systemic understanding of patterns of practice in classrooms where teachers are trying to enact reform" (Spillane & Zeuli, 1999, p. 20). Gage (1978), Shulman (1983), and others have argued convincingly for the value of case studies as existence proofs, illustrating what can be accomplished rather than documenting what is typically the case.

One major virtue of a case study is its ability to evoke images of the possible. . . . It is often the goal of policy to pursue the possible, not only to support the probable or frequent. The well-crafted case instantiates the possible, not only documenting that it can be done, but also laying out at least one detailed example of how it was organized, developed, and pursued. For the practitioner concerned with process, the operational detail of case studies can be more helpful than the more confidently generalizable virtue of a quantitative analysis of many cases. (Shulman, 1983, p. 495)

Yet, researchers have noted that case studies alone have limitations for investigating state and district reform efforts, and they have suggested the potential value of conducting quantitative investigations in addition. For example, after offering insights about the logic of systemic reforms, the avenues by which they may reach classrooms, and their impact on classroom practice, based on a review of

qualitative investigations of large-scale systemic reform initiatives in mathematics and science, Knapp (1997) suggested:

Obviously, case studies and qualitative findings such as those reviewed here give little indication of system-wide trends and tendencies, and even though intelligent guesses can be made in some instances, there is a clear need for large-sample research that can locate case study patterns in a larger, system-wide context. (p. 257)

Similarly, those who study teaching using quantitative strategies have recognized the value of complementing their efforts with more qualitative investigations. As Linn (1990) noted:

The person using quantitative methods must make many qualitative decisions regarding the questions to pose, the design to implement, the measures to use, the analytical procedures to employ, and the interpretations to stress. (p. 1)

A few researchers have followed these suggestions and used multiple methods to study school reform, but often these approaches have been sequential rather than integrated. One common approach is to use data from surveys as a basis for sampling teachers for more in-depth observation and interview. For example, Spillane and Zeuli (1999) used a large-scale teacher survey drawn from the Third International Mathematics and Science Study (TIMSS) to describe mathematics teaching in nine districts and to select teachers who reported teaching “in ways that resonated with the reforms” (p. 3). This approach to sampling allowed them to focus their observations and interviews on classrooms where practice was more consistent with the goals of the reformers.

Similarly, Schorr and Firestone (2001) collected survey data about mathematics and science teaching practices from 245 fourth-grade teachers in New Jersey. Using the results of the survey, they selected a subsample of 23 teachers who scored at the extremes on scales of “inquiry-oriented” practices and direct instructional practices. These teachers were observed on two occasions and interviewed in order to help the researchers “clarify teacher’s responses” to the state policy.

Smith (1997) provides an example, albeit a rare one, of a more integrated use of survey and case studies. In studying the implementation of the Arizona Student Assessment Program (ASAP), Smith and her colleagues used both techniques in complementary ways. They conducted four interrelated studies: year-long case studies of four schools, a follow-up focus group inquiry at the case study sites,

statewide surveys of educators, and supporting studies of assessment plans and policymakers' beliefs. All these data were combined during the analysis to develop assertions about the Arizona program. For example, the authors argue that the reform intention conflicted with the accountability intention of the testing program, and this conflict impeded coherent action on the part of schools. The evidence for this assertion is drawn from interviews conducted as part of the case studies, both closed and open-ended responses to the statewide survey, the follow-up focus group study and a supporting study (Smith, 1997, pp. 82–91).

Borman and colleagues plan to perform similar integrated analyses of survey and case study data as part of their study of the NSF Urban Systemic Initiative, but analyses reported to date rely on single data sources (Borman & Lee, 2001; McCourt, Boydston, Borman, Kersaint, & Lee, 2001).

Smith (1997) argues that the key to achieving maximum benefit from using multiple approaches is the integration of the data they provide:

Qualitative approaches are necessary to understand how educators are defining and coming to terms with the reforms and to look closely at what they are actually doing about it. At the same time, it is helpful to be able, through survey techniques, to gauge the range of beliefs and practices subsequent to the implementation of the reform. Each contains certain assumptions, and each supports different kinds of inferences. The strength of the analysis is the linking of data from the whole. (p. 9)

There were many instances in our project in which being able to consider data from both survey and case study sources enhanced our understanding of the impact of the Washington reform on schools and classrooms. However, we were also frustrated that the data did not permit us to address all the questions we had posed. The case studies did not always illuminate intriguing results from the surveys, and the surveys sometimes failed to place case study insights in a larger context. Some of these shortcomings reflect the inherent incompleteness of research. However, there were several cases in which we felt, in retrospect, that we could have improved the study by doing things differently. This paper looks closely at the complementarity of our methods, and considers how they might be used more effectively to study school reform in the future. We try to answer the question of how surveys and case studies can best be used in combination in the study of school reform to maximize the usefulness of the data they provide.

Methods

Our efforts to combine survey and case study methods in the study of assessment-based school reform began in 1996 with studies of the impact of the Kentucky Instructional Results Information System (KIRIS). The Kentucky reform began in the early 1990s, and it was relatively “mature” by the time we began our research. Our goal was to use the surveys to describe the larger landscape of the state reform effort and the case studies to reveal more detailed information about selected locations (Borko & Elliott, 1999; Stecher, Barron, Kaganoff, & Goodwin, 1998; Stecher & Barron, 1999; Wolf, Borko, Elliott, & McIver, 2000; Wolf & McIver, 1999).

In 1998 we shifted our focus to the state of Washington, which was just initiating a standards-based reform effort and thus provided an opportunity to study the impact of high-stakes state assessments at an earlier stage of implementation. Again, we combined the qualitative and quantitative strengths of research teams from our two institutions to study this reform through surveys and case studies.

This paper focuses on our study of the Washington education reform because our research methods were better coordinated in Washington than in Kentucky. Our experiences in Kentucky helped us improve our joint design and data collection efforts in ways we hoped would increase the benefits of using both qualitative and quantitative approaches. In addition, a preliminary analysis suggested that although consideration of data from both states might reveal different substantive themes, the advantages and limitations of coordinating the two methods are similar in both states. We are working on a separate summary of our substantive findings regarding the impact of the Washington education reform on schools and classrooms. Here we focus on the integration of the methods.

To promote complementarity between the surveys and case studies, we designed the data collection strategies collaboratively. The survey plan was to begin with a broad focus the first year, and narrow our attention in the second year. We would focus, in that second year, on factors that appeared most salient, such as formal and informal support for reform, patterns of assignments, and judgments about the quality of student work products. The case study plan was to focus on a carefully selected set of exemplary schools. This purposeful sample of schools was chosen through discussions with a variety of people familiar with the reform agenda

in the state, including the superintendent and members of her staff, personnel at regional and district offices, and university faculty members, as well as visits to a set of schools whose names came up repeatedly in these conversations. We also attempted to increase the informal integration of the two efforts. For example, we conducted shared site visits, in which researchers from the survey team and the case study team visited districts and schools together (during both the site selection and data collection phases of the project). The conversations that took place as we gathered and processed information together helped each team better understand the perspectives of the other. (For more detailed information about the survey methods, see Stecher, Barron, Chun, & Ross, 2000; for more detailed information about the case study methods, see Borko, Wolf, Simone, & Uchiyama, 2001).

We coordinated the focus of survey and case study instruments to provide comparable data across the two types of investigations. We began with document reviews and interviews with state and district administrative staff to become familiar with the reform goals, supports, and incentives. These background data were used to design surveys that were responsive to the local context and the manner in which the reform was implemented at the district and school levels. The reviews and the interviews also were used to inform the development of observation guides and interview protocols for the case studies to focus on relevant changes in practice, perceived changes in the character and quality of student work, the nature and effectiveness of available supports, and desired additional resources. The two teams of researchers met to talk about the reforms and agree on areas of emphasis, as well as shared language to use in the survey and case study investigations. Thus, several common foci were built into survey and case study data collection instruments.

The analysis of these data proceeded in stages. Initially, the surveys and case studies were analyzed separately. Survey results were examined using traditional quantitative analytic techniques. We computed frequency distributions for items with fixed response options, and we computed mean values and standard deviations for questions requiring a numeric response. When looking at differences in responses among various behaviors, we focused more on practical significance than on statistical significance (e.g., doing something daily compared with doing it only once a week, or having 20% more teachers hold one opinion than another). In a few cases we used more complex quantitative techniques to review the data. For example, we conducted a factor analysis to identify underlying traits in the survey

responses, and we used regression models to look for associations among these traits and school test scores controlling for student demographic characteristics. Comprehensive reports of the survey results were published and shared with educators in Washington and were presented at the annual meeting of the American Educational Research Association (Stecher et al., 2001).

Similarly, the case study analyses used traditional qualitative techniques. During each of three 2-day visits to each site, researchers observed writing and mathematics instruction; conducted semi-structured interviews with teachers, students, and the school principal; and collected artifacts relating to instruction and school programs. The interviews focused on writing and mathematics practices, professional development opportunities and experiences, and knowledge and beliefs about the Washington education reform. At each site researchers also interviewed district personnel about the impact of the reform. After each observation, the notes were expanded into detailed field notes. The interviews were audio taped, and these tapes were transcribed into printed text. The analysis began by summarizing the field notes in condensed “cover sheets,” using a set of categories derived from the research questions and themes that emerged from an initial reading of the data. The transcripts and any artifacts collected during the site visits were coded in a similar manner, ensuring that the coding reflected both the conceptual framework used in formulating the interviews and any supplemental understanding that emerged from reading the transcripts. Following these efforts, analytic case study narratives were written for each school based on all the information and understandings that were available. These became the “raw data” for cross-site analyses. These results were presented at the annual meeting of the American Educational Research Association (Borko et al., 2001).

As a first step in considering how to integrate findings from the surveys and case studies, members of each research team read the comprehensive analysis reports prepared by the other team. After reviewing the survey and case study data separately, we looked at the data jointly. We compared results from the survey with insights gained from the case studies to try to build a more complete understanding of the impact of the reform. Impressions derived from one source were tested against the data from the other to look for supporting or contradictory evidence. There were a number of substantive themes we could explore with data from both survey and case studies. (Despite our joint planning efforts, there were also many

themes that emerged from one source for which we found little or no relevant information from the other source, a point we return to in the discussion.)

We selected six themes for inclusion in this paper, not because they were the most important from the perspective of assessment reform policy but because they provided examples of ways in which the survey and case study data were complementary. We anticipated that the two methods would complement each other in a number of specific ways. First, the case studies would provide specific information to describe a more general observation, helping to illustrate “what” occurs in greater depth and with more clarity. Second, case studies would reveal explanatory examples that illuminate “why” relationships exist or “how” one factor influences another. Third, surveys would provide information about the prevalence of patterns found in case study sites and, perhaps, help us to understand what makes those sites exemplary.

In the following section we draw upon the survey and case study findings to examine whether and how the two methods complemented each other in our project. Our preliminary analysis suggested that we could explore all of the anticipated relationships between the methods by limiting our analysis to a closer look at the three exemplary elementary schools, rather than attempting to incorporate data from both the elementary and middle schools. Thus our examples, which highlight both situations in which expectations for analytic complementarity were met and other situations in which they were not, focus on the 1999 and 2000 surveys and the three exemplary elementary schools (Beacon, Emerald, and Vista—all names of schools and school personnel are pseudonyms).

Results

The surveys and case studies were complementary in the ways we anticipated. We present some specific examples below. While these illustrations demonstrate that the methods can support each other, they do not indicate that they complemented each other to the degree that we think is possible. In fact, we also found numerous occasions in which the case studies provided little or no information to clarify a finding from the surveys, and others in which there was no information on the surveys regarding the prevalence of particular patterns revealed at the case study sites. Further, in some ways our design was less than optimal in terms of maximizing the integration of the two methods.

We begin with examples of useful synthesis offered by the two methods, and then address the question of how our design might be modified to better integrate the two methods. We present these examples organized around six reform themes. The first four themes focus on broad patterns of curriculum and instruction: curriculum alignment, changes in curriculum and instruction in mathematics, writing instruction with an emphasis on genre, and the reallocation of instructional time. The fifth and sixth themes use the Washington Assessment of Student Learning (WASL) as a starting point, focusing on the impact of the assessment on teachers and their practices, specifically, the relative reported impact of the WASL and Essential Academic Learning Requirements (EALRs), and teachers' instructional practices. The descriptions include instances in which the combination of surveys and case studies enhance our understanding of "what," "why," "how," and "how much."

Curriculum Alignment

Curriculum alignment is often an early step in a systemic reform effort (Knapp, 1997). This was certainly the case at our survey and case study schools. The vast majority of teachers surveyed in both spring 1999 and spring 2000 reported that they understood the process of aligning curriculum and instruction with the EALRs. The percentage of fourth-grade teachers who said they understood it well or very well increased from 78% in 1999 to 85% in 2000. The increase was greater among seventh-grade teachers than fourth-grade teachers. It rose from about two thirds of seventh-grade writing and mathematics teachers in 1999 to almost 90% in 2000. There was also an increase in the percentage of fourth- and seventh-grade teachers who said their professional development focused a moderate amount or a great deal on aligning curriculum and instruction with the EALRs. Among fourth-grade teachers, the percentage increased from 42% in 1999 to 65% in 2000. There was a comparable increase in the percentage of seventh-grade writing teachers who reported a moderate or a great deal of emphasis on alignment in their professional development (from 47% in 1999 to 73% in 2000). In contrast, 61% of seventh-grade mathematics teachers reported that their professional development emphasized alignment in 1999, a figure that grew only slightly to 64% in 2000. However, only about one half of the teachers surveyed reported that their curriculum was "very well aligned" with the EALRs in reading, writing, and mathematics in 2000. The highest percentage (62%) occurred among seventh-grade writing teachers. Between

42% and 45% of fourth-grade teachers said their reading, mathematics, and writing curricula were very well aligned with the EALRs.

At the time of our study, all three case study elementary schools had made significant progress in aligning their curriculum with the EALRs and WASL. Their experiences provided some examples of the kinds of practices that underlie the survey responses. By the time of our first visit, for example, Beacon Elementary School had been conducting curriculum alignment meetings by grade level for several years, intentionally aimed at meeting the challenges of the WASL. More recently, staff had been meeting across grade levels to gain a broader understanding of how each grade's curriculum flows into the next, and then to fill in the gaps. At these meetings teachers in each grade report to the rest of the staff how their students will meet the "targets," or the Essential Learnings. They discuss and refine strategies to make the transition from one grade to the next as seamless as possible. "In mathematics," Ms. Watson (the fourth-grade mathematics teacher whom we observed) explained, "we all come together with what we have, and we ask ourselves, 'Where are we going? Look at the EALRs. What important pieces do we really need to hit?'" (T00S).

Alignment activities were also extensive at Vista Elementary School, and were a major component of the school's efforts to improve students' opportunities to learn. Mathematics was the first content area to be addressed. At each grade level, teachers "lined up" lessons in their textbook with the content and processes tested on the WASL. By doing this, "It became really clear where the holes were" (T00S). And there were "some significant holes" (P00S), particularly in the areas of the WASL where students' scores were lowest—probability, reasoning, and communication. In contrast, teachers had not yet aligned their reading curriculum with the EALRs. They were in the process of adopting a new reading series during the 1999–2000 academic year and planned to begin the work of aligning that series with the EALRs during the following summer.

The fact that alignment efforts were still underway at the exemplary schools is not surprising, since the Washington reform is recent. This timing also helps explain the survey pattern where, although the vast majority of teachers understood the alignment process well, only about half reported that their schools' curricula were well aligned with the EALRs in reading, writing, and mathematics.

Change in Curriculum and Instruction Within Mathematics

Teachers modified their classroom practices in mathematics and writing in ways that were compatible with their work on curriculum alignment. In mathematics, there were changes in the frequency with which teachers addressed each of the five content areas and each of the 11 process skills identified specifically in the EALRs. Focusing on the areas of greatest change, about 40% of the teachers in fourth grade and about one third of the teachers in seventh grade reported increasing their coverage of probability and statistics and of algebraic sense (see Table 1).

The mathematical processes for which the greatest percentages of teachers reported increasing coverage included gathering information, analyzing information, representing and sharing information, and relating concepts within mathematics (see Table 2).

The case study of Vista provides a more detailed illustration of the impact of the reform on classroom curricular and instructional decisions in mathematics. As one component of their alignment effort, Vista’s teachers determined which lessons in the mathematics textbook to teach and which to omit (because there were more lessons in the text than instructional days in the school year). Then, based on the holes they had identified in the curriculum, they also identified materials that they could use to supplement their text-based instruction, such as state-provided tool kits and timed tests, as well as the Every Day Counts and Problem of the Day supplementary materials. For example, Ms. Thompson, the principal, required that

Table 1
Frequency of Coverage and Increase in Coverage of Mathematics Content EALRs in 2000 (percentage of teachers)

Mathematics content EALRs	Daily or weekly coverage of EALR		Increased coverage since 1999	
	Grade 4	Grade 7	Grade 4	Grade 7
1.1 Number sense	79	68	33	24
1.2 Algebraic sense	37	57	40	31
1.3 Measurement	27	23	25	16
1.4 Geometric sense	27	20	34	24
1.5 Probability and statistics	22	14	43	33

Note. EARLs = Essential Academic Learning Requirements.

Table 2

Frequency of Coverage and Increase in Coverage of Mathematical Process EALRs in 2000
(percentage of teachers)

Mathematics process EALRs	Daily or weekly coverage of EALR		Increased coverage since 1999	
	Grade 4	Grade 7	Grade 4	Grade 7
3.1 Analyze information	75	70	41	44
3.3 Draw conclusions and verify results	68	56	39	27
2.3 Construct solutions	68	59	29	32
5.3 Relate concepts to real life	65	82	42	34
4.2 Organize and interpret information	62	48	45	32
2.1 Investigate situations	62	58	35	28
5.1 Relate concepts within math	62	66	49	35
2.2 Formulate questions	56	49	46	40
4.3 Represent and share information	54	32	55	45
4.1 Gather information	51	43	50	31
5.2 Relate concepts to other disciplines	43	43	45	27

Note. EARLs = Essential Academic Learning Requirements.

all classes do a Problem of the Day because “our kids were missing out on some real opportunities to think at higher levels, and to analyze problems in your head and do them in your head” (P00S).

Ms. Erikson, the fourth-grade teacher who participated in our study, closely followed the textbook for the majority of her mathematics instruction. She characterized a typical lesson as “fairly traditional, since we use the book. We usually start off by correcting the previous day’s lesson and talking about any difficulties that there were. . . . Then basically I do some sort of a set for the lesson. . . . Usually there’s . . . guided practice . . . with questioning and having students explain their thinking as they work through the problems. . . . Then student independent work, and I roam around the room at that time” (T99S). Although her curriculum was aligned with the EALRs and she did incorporate student written and oral explanations into her lessons, to a large extent Ms. Erikson’s instructional practice had not been affected by the Washington reform. The exception to this pattern was her use of supplementary materials. As one aspect of her effort to meet the expectations of the reform, Ms. Erikson began using tool kit activities systematically during the 1998–1999 school year. She explained that she depended on the framework and the tool kit together, “because I think our book, even though

it aligns with the Essential Learnings, is missing a lot of big pieces. I don't think it provides enough higher level thinking skills" (T99S). During tool kit activities, students were engaged in solving richer mathematical problems and communicating their mathematical thinking to a greater extent than during textbook-based lessons.

Ms. Wright's fourth-grade mathematics program at Emerald Elementary School is more reform-oriented than Ms. Erikson's. Although she weaves lessons on computation and skills throughout the program (e.g., students take a timed facts test each day), activities that focus on conceptual understanding have a more prominent role in her mathematics class. Like teachers at Vista, Ms. Wright assigns a Problem of the Day; this problem requires students to use problem-solving and thinking skills, and to explain their thinking in pictures, numbers, and/or words. In addition, a typical class period includes mental mathematics activities, a formal lesson on a concept, student independent problem-solving activities (often done in small groups), and writing about mathematics.

In contrast to the situation at Vista, mathematics instruction at Emerald began to change prior to the onset of WASL testing, in large part because of a grant received by the district, which provided professional development in reform-based mathematics for all teachers in the district. In the fall of 1999, Ms. Wright commented, "I believe Marilyn Burns has been in the district for the past five years" (T99F). She explained, "Not only do we have a week-long institute with her every summer, but we have one-day classes throughout the year that we can choose to go to" (T99F). On the other hand, although Vista began its curriculum alignment efforts with mathematics, at the time of our study its professional development activities had focused almost exclusively on reading and writing. These differences between the two schools may help explain the variety in instructional content and practices revealed by the surveys.

Writing Instruction With an Emphasis on Genre

Survey responses revealed that most teachers addressed most of the writing EALRs on a weekly or daily basis (see Table 3). For example, according to the 2000 surveys, over 80% of fourth- and seventh-grade teachers addressed the application of writing conventions on a daily or weekly basis, and over 50% addressed each phase of the writing process at least weekly. Attention to features of writing related to genre was somewhat less frequent in both grade levels. However, there was

Table 3

Frequency of Coverage and Increase in Coverage of All Writing EALRs in 2000 (percentage of teachers)

Writing EALRs	Daily or weekly coverage of EALR		Increased coverage since 1999	
	Grade 4	Grade 7	Grade 4	Grade 7
Application of writing conventions	84	84	34	40
Writing process: draft	71	70	39	41
Writing process: edit	65	70	38	37
Writing process: pre-write	65	69	39	40
Writing process: revise	63	68	40	44
Development of concept and design	47	36	56	42
Seek and offer feedback to other students	46	38	37	33
Genre: write for different purposes	45	67	45	57
Genre: style appropriate to audience and purpose	41	54	38	42
Genre: write for different audiences	36	45	41	42
Genre: write in a variety of forms	35	64	41	53
Writing process: publish	35	49	29	36
Assessment of students' own strengths and needs for improvement	33	37	39	34
Write for career applications	12	18	16	18

Note. EARLs = Essential Academic Learning Requirements.

considerable change occurring in the teaching of writing. One third or more of the teachers in both grade levels reported increasing their coverage of each of the writing EALRs between 1999 and 2000. The features related to genre, including writing for different audiences, writing in a variety of forms, using style appropriate to audience and purpose, and writing for different purposes, were among those that received increased coverage from the greatest percentage of teachers. Between 35% and 45% of fourth-grade teachers and between 45% and 67% of seventh-grade writing teachers reported that they increased their coverage of these EALRs from the previous year.

A closer look at the types of writing assignments shows differences between the grade levels, but a growing emphasis on the types of writing that are tested as part of WASL (see Table 4). Three of the four writing genres were covered at least once a month by 64% or more of the fourth-grade teachers. Persuasive writing appears to be less common in fourth grade. In addition, half of the fourth-grade teachers reported that they had increased their coverage of expository writing since

1999. Thus, fourth-grade teachers appeared to be putting greater emphasis on narrative and expository writing, which are the focus of the fourth-grade WASL. Seventh-grade teachers appeared to be making changes to place more emphasis on persuasive and expository writing (which are the focus of WASL in seventh grade). While seventh-grade teachers were somewhat more even in their coverage of writing genres than fourth-grade teachers, persuasive and expository writing were covered at least monthly by the greatest percentage of teachers and had seen their coverage increasing by the greatest percentage of teachers.

Descriptions of typical writing instruction in target classrooms at the case study schools provide a detailed picture of teachers’ attention to writing conventions, the writing process, and genre. For example, at Vista, Ms. Erikson organizes her writing program around “working through” various genres, having students engage in each step of the writing process as they study each genre. To prepare for the WASL, she teaches four genres: narrative, procedural, recount, and expository. Once all four are introduced, “we continue to go back through them the whole year, . . . to spiral and get better at each one as we go through” (T99F). This organization around genres was apparent to us even before we observed Ms. Erikson’s teaching—in our conversations with her, the resources she used for writing instruction, and materials displayed on the classroom walls. In an initial interview, she explained, “We put the different forms up on the walls as we go, and list the different elements of their frameworks. . . . And as we go through them, I’ll throw out a topic and ask . . . ‘if I wanted to write something more about this topic, and have it be truthful and factual, how would I best do that?’ . . . We talk about why a report would be better to tell somebody about the life cycle of the salmon,

Table 4
Frequency of Coverage and Increase in Coverage of Writing Genres in 2000 (percentage of teachers)

Writing genres	Daily or weekly coverage of EALR		Increased coverage since 1999	
	Grade 4	Grade 7	Grade 4	Grade 7
Narrative	76	44	37	16
Persuasive	29	59	32	50
Expository	64	63	50	43
Descriptive	64	57	35	19

Note. EARLs = Essential Academic Learning Requirements.

rather than some sort of fantasy narrative. Or why it would be better than a procedural writing” (T99S).

Prominent among Ms. Erikson’s instructional materials are the students’ writing folders. These folders illustrate how she uses a writing process approach to help students become proficient with each genre. Each folder we looked through contained samples of student work in various genres, for each step of the writing process, as well as several writing resources (e.g., a pamphlet titled “My Writing,” another titled “Quick-Word: Handbook for Everyday Writers,” and a sheet titled “Checklist for My Writing”). Each folder also contained a writing sample on the topic “My Favorite Place.” A tool kit activity on the writing process provided the guidelines for this piece. Ms. Erikson explained, “It’s one that I use toward the beginning of the year. . . . There’s a whole packet—a checklist that they go through for pre-writing; a checklist for their first draft, second draft, and even . . . third draft; and revision and editing” (T99S).

Ms. Erikson teaches writing conventions through Daily Oral Language (DOL) activities. Each day, DOL includes a number of exercises that address spelling, punctuation, parts of speech, and other writing activities. For example, one typical activity involves correcting errors in sentences written on the chalkboard; another is sentence dictation. The variety in DOL activities helps ensure that all students’ needs for learning writing conventions are addressed, while enabling Ms. Erikson to focus on other aspects of good writing, such as “being clear” and “style—trying to come up with better ways of saying things and more interesting words” (T99S) during other components of the writing program.

Like Ms. Erikson, Ms. Wright weaves instruction on the writing process into her teaching of the various genres, focusing on each step of the process from prewriting to publishing. She explained, “I taught them a couple of ways to pre-write. I said sometimes writers do pre-writing in their heads. Some people have that ability. Others have the ability to do it on paper and to web it out. So you choose whatever pre-writing you want” (T99F). Although she teaches her students revising and editing, she does not have them take every assignment through the final steps of the writing process. For example, if they have been working for several weeks on a particular genre and have written rough drafts of multiple pieces in that genre, she will tell them to “pick one and take it through the writing process. . . . I let them pick one of the ones that they feel more ownership with. . . . They have to revise and edit

themselves. And then they get with a partner and the partner revises and edits their work” (T99S).

Ms. Wright’s approach to instruction in the genres differed from the approaches of Ms. Erikson and Ms. Alexander (the fourth-grade language arts teacher we observed at Beacon) in a way that may help to explain one of the survey findings about writing practices. During the 1999–2000 school year, Ms. Wright did not give equal attention to all genres addressed in the EALRs. Rather, she focused more closely on narrative and expository. She explained, “This year I really wanted to focus on narrative and expository and really hit them hard, because those were the ones tested on the WASL. I spent the first part of the year teaching narrative and the second part, before the WASL, teaching expository. In between those, I hit the different genres” (T00S–W 187). In 1999 OSPI decided that the fourth-grade writing prompts would require narrative and expository writing, while the seventh-grade prompts would require persuasive and expository writing. If other teachers in the state made decisions similar to Ms. Wright’s, this fact may explain the survey finding that fourth-grade teachers focused more on narrative and expository writing than on the other genres. Our case study data also suggest that, at least in this specific arena, some exemplary sites differ from typical schools in that they do not limit their instructional programs to the more narrow requirements of the WASL, but instead focus on the broader set of learnings encompassed by the EALRs.

Reallocation of Instructional Time

The Washington Education Reform has led fourth-grade teachers to reallocate their instructional time across subject areas. At present, fourth-grade teachers spend 5 hours per week on reading and mathematics; 4 hours on writing; 2 hours each on communication, social studies, and science; and 1 hour on arts and health and fitness. Other subjects account for an hour, as well. This pattern of time allocation represents a shift in the use of instructional time during the period we were studying. More time has been devoted to the subjects that are currently tested as part of WASL—reading, writing, mathematics, and communication. Time has been taken away from the nontested subjects—social studies, science, arts, and health and fitness. Figure 1 shows the percentage of teachers reporting changes in coverage (either an increase or a decrease) for each of the subjects covered by the EALRs. For example, in 2000, 59% of fourth-grade teachers reported increasing the time they spent on writing and mathematics, while 40% to 50% reported decreasing the time

they devoted to social studies, science, and art. Similar results were reported in 1999, which suggests reallocation is continuing.

The case studies help us understand time reallocation from the teachers' perspective. At all three case study schools, teachers talked about the importance of making decisions regarding the allocation of instructional time to conform to the EALRs. They used words like "intentional" and "focused" to characterize these decisions. At Beacon, Ms. Alexander commented, "The EALRs and the state guidelines are helpful, because it's one thing to teach what you're passionate about or what you care about or what you think is important. But that might be studying frogs for two years, and maybe that's not the best use of the students' time. So I think the EALRs help us be more intentional about what we're teaching" (T99F). Similarly, Ms. Wright (at Emerald) shared her belief that "the EALRs and the WASL allow for more teacher knowledge of what needs to be taught." She saw them as a

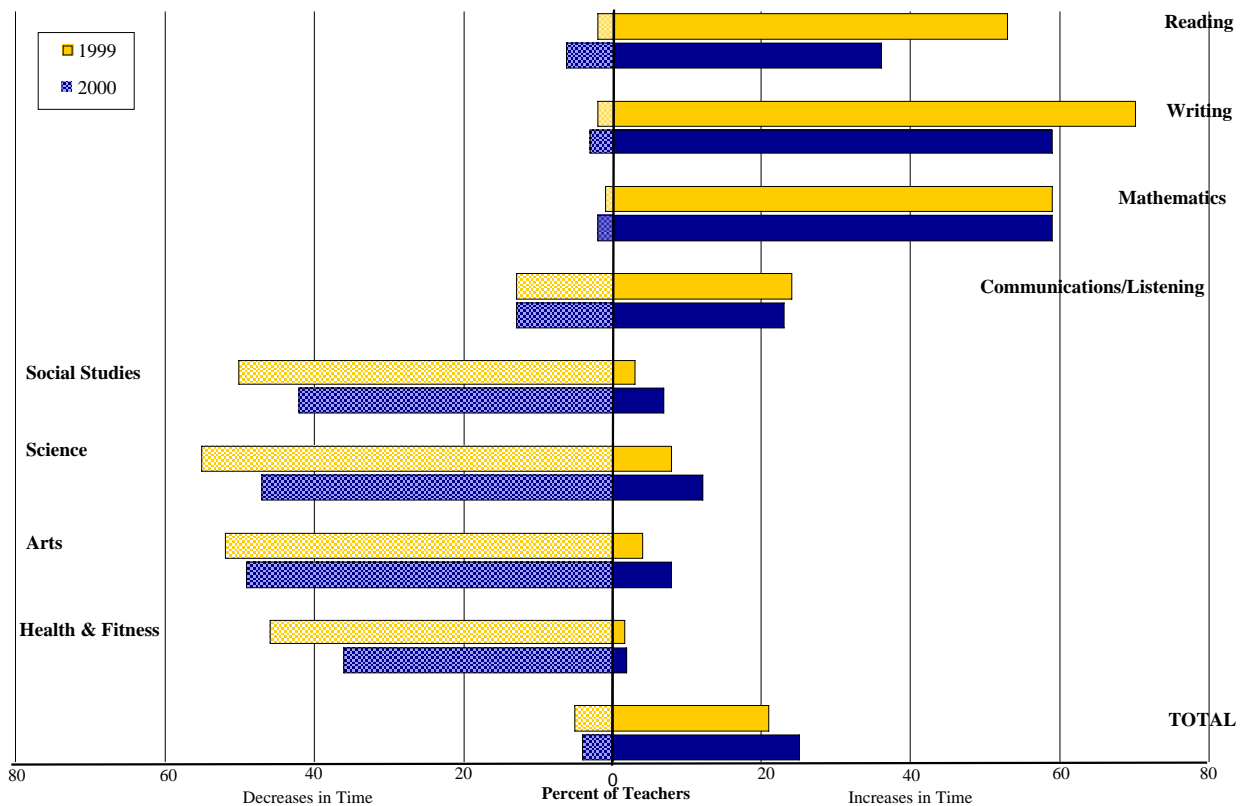


Figure 1. Washington Assessment of Student Learning (WASL) scores for students in Grades 4, 7, and 10.

“welcome guide,” explaining that “I’ve welcomed the EALRs because they help me as a teacher know what I need to teach. The WASL is not everything that you need to teach, but it gives you a good foundation, a good core to jump to.” At Vista, Ms. Erikson gave a specific example of how the EALRs and WASL had helped her to be “more intentional and explicit in my teaching” (T00S) and to take “the fluff” out of her curriculum. She replaced several of her favorite activities with other activities that were more closely aligned with the Essential Learnings. As she explained, “I loved doing the Iditarod unit. I haven’t done it for the last couple of years because it just doesn’t flow exactly with what we’re doing” (T00S).

The principals at all three schools promoted decisions that brought curriculum and instruction into closer alignment with the goals of the Washington reform. Ms. Thompson, the principal at Vista, explained that she was “very interested in making sure that we don’t waste time teaching things that don’t align with the EALRs” (P99S). Every teacher at Vista was required to identify the Essential Learnings addressed in each lesson, and the administration checked lesson plans weekly. The principal and associate principal used a “cheat sheet for the Essential Learnings” when observing teachers, and they referred to these sheets in post-observation conferences to highlight ways in which the lesson was and was not aligned with the EALRs (P99S).

The case study teachers and principals did not discuss the reallocation of time across subject areas in their conversations with us. This silence may be the result of the explicit attention to mathematics and language arts in our observation schedules and interview protocols. Thus, the case study data do not provide further insights regarding this important phenomenon revealed by the surveys.

Looking across the several aspects of curriculum and instruction we have addressed, we see clear evidence that Washington’s reform is in transition. Patterns of partial implementation and changes over time are prevalent in the survey data. The ongoing change efforts that principals and teachers described and researchers observed at the case study schools provide concrete illustrations of the variety of ways in which this reform is unfolding. One of the most dramatic examples of this variety is the different timelines for mathematics reform at Emerald and Vista. As we noted earlier, because of a mathematics grant received by the district, teachers at Emerald began participating in professional development in mathematics reform several years before the WASL was given. In contrast, although they had participated in extensive professional development in reading and writing for a

number of years prior to WASL testing, teachers at Vista had received virtually no professional development in mathematics by the time of our study. As Ms. Thompson explained, “We certainly haven’t focused on mathematics as much as we have on reading, writing, and language acquisition, because we really believe that those three are the foundation. You’ve got to have those before you can do the others” (P99S). Not surprisingly, the mathematics instruction we observed at Emerald matched the content and processes specified in the EALRs to a much greater extent than the instruction we observed at Vista.

Reported Impact of the WASL and EALRs

Principals and teachers appeared to be focused on the WASL more than the EALRs. This impression comes from responses to a number of different survey questions. Almost all principals and teachers who responded to the surveys said they felt a moderate amount or a great deal of pressure to have students do well on the WASL. As illustrated in Figure 1, teachers shifted class time to subjects currently being tested, even though the state has adopted EALRs in several other subjects for which the tests are not yet developed.

We presented teachers with two contrasting viewpoints on addressing the EALRs and WASL, and asked them to identify their own approach relative to these two. The first point of view focused on the standards: “I teach the EALRs, and I don’t bother with WASL preparation at all. If students master the EALRs, they will do well on the WASL.” The contrasting viewpoint focused on the test: “I teach to the WASL, and I make sure my students practice the kinds of questions they will encounter when they take the test. It is important for students to master the material on the WASL.” Teachers were asked to rate their own approach as being “just like” one of these two extremes, “somewhat like” one of these two, or not like either of them. Two thirds of teachers identified their teaching as more like “teach(ing) to the WASL,” than “teach(ing) the EALRs.” Principals responded similarly when asked about the approach they encouraged at their schools.

Case study findings on schools’ use of WASL scores help us to understand why the surveys indicate a greater impact of the WASL than the EALRs on teachers’ instructional practices. Principals at all three elementary schools conducted detailed analyses of WASL scores, on a class-by-class basis, for their teachers. Teachers, in turn, used this information on WASL performance to guide their curriculum and instruction in very concrete ways.

At Beacon, because of her background in assessment, Ms. Powers leads the school in using the WASL to guide decisions about curriculum and instruction. Each year, as soon as the scores arrive, Ms. Powers takes them home, enters all the data on her computer, and carefully studies her school's performance. She makes graphs for each teacher, displaying student scores from the current and previous years, growth rate, and how close each is to the goals. She also runs off a list of scores for all the schools in the district, to see where Beacon falls.

Ms. Watson commented, "We had a staff meeting to discuss the scores. Ms. Powers showed the test scores and did some comparisons with previous years, and with our goals for the current year. So, not only are we always looking at how we compared to the past, but where we're going in the future to make improvements" (T99F). Beacon teachers used this information to plan their next curricular moves. Recently, for example, they focused on communication skills. "We looked at the things the students were struggling with. One of the biggest was communicating their ideas and showing their understanding . . . so we continue to work on communication and understanding," said Ms. Alexander (T99F).

One of the central professional development activities at Emerald is a schoolwide staff retreat held just before the beginning of the school year. In August 1999, the major agenda items at the retreat were to analyze WASL scores and develop goals for the year based on those scores. Ms. Glen presented the test results to the teachers, "pinpointed the areas from the data that showed our strengths and areas of weakness," and presented broad building-level goals in reading and mathematics that the school's Learning Team had identified based on their initial discussions of the WASL scores. As Ms. Glen explained, "We put the goals up on butcher paper and said, 'OK, it looks like these are the things we are going to need to focus on.'" Then, "the staff broke into grade-level groups to discuss, 'Do these goals feel right? Anything we want to add?'" The Learning Team used the results of these discussions to formulate 15 more specific goals in reading and 10 in mathematics, which became the focus for the building during the 1999–2000 school year (P99F).

Vista teachers also found the concrete information about student performance on specific learning goals to be valuable in planning their curriculum and instruction. Ms. Thompson, like Ms. Powers, conducted a detailed analysis of each fourth-grade class's performance on the WASL, comparing percentages of students who achieved a passing score in each of the assessment areas to percentages for the

school as a whole. She also looked at differences in the school's scores across years. Based on these analyses, the fourth- and fifth-grade teachers selected areas of schoolwide instructional emphasis for the 1999–2000 academic year; these included listening, nonfiction, and five areas of mathematics for which scores were particularly low—reasoning, making connections, probability and statistics, number sense, and algebraic sense. Thus, as Ms. Thompson explained, WASL scores “raised our awareness level in terms of where we need to put our energies” (P00S).

In all three cases, the results from the WASL were used as a basis for instructional planning. The test results provided concrete, curriculum-related information about student performance, specific to the school, which could be studied and acted upon. Schools rarely have access to information that can be used for curriculum reform. For example, nothing comparable provides comprehensive evidence of student mastery of the EALRs. The uniqueness of the WASL results as tools for reform may help explain the survey finding about apparent greater attention to the WASL than the EALRs.

Test Preparation

Teachers were asked to indicate the frequency with which they used eight different methods to prepare students for the WASL tests, including discussing the EALRs, reviewing general test-taking strategies, practicing with released WASL items, and scoring work using the WASL rubrics. In mathematics, 40% or more of the teachers had students practice using released WASL items, discussed responses to WASL items to illustrate levels of performance, and used open-ended questions as a part of classroom work at least once each week. In writing, 50% or more of the teachers taught rubric-based approaches to writing and used open-ended questions in class at least weekly to help students do well on the WASL. There was considerable range in how frequently the various test preparation approaches we asked about were used. For example, about one half of the teachers said they used open-ended questions (short-answer and extended-response) in class work at least weekly, whereas only about 10% used material from the assessment tool kits at least weekly.

There was also considerable variation in the time devoted to test preparation during the year. Figure 2 shows that the percentage of fourth-grade teachers who spent 4 or more hours per week preparing students for the WASL in mathematics increased from 6% in November to 35% in April. Among seventh-grade

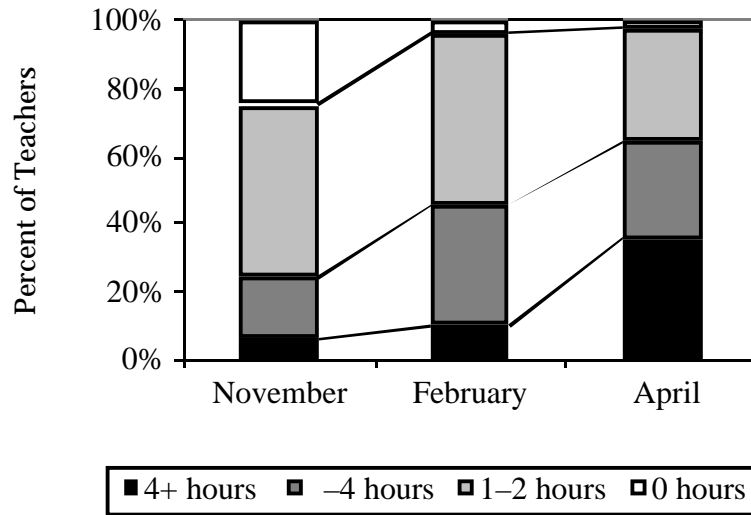


Figure 2. Hours per week spent in fourth-grade classrooms preparing for the WASL test in mathematics in 2000.

mathematics teachers, the percentage who spent 4 or more hours per week preparing for the WASL grew from 4% in November to 21% in April. In writing, the percentage of fourth-grade teachers who spent 4 or more hours per week grew from 5% in November to 18% in April, while among seventh-grade writing teachers the percentage grew from 6% to 36%.

Teachers at all three case study schools are very deliberate in helping students prepare for the writing and mathematics portions of the WASL. Ms. Alexander has them write to the sample prompts provided by the state, and she scores their pieces with the students' help, using the state-provided assessment criteria. Then she gets together with groups of students to talk about what worked for them as they were writing and what strategies they used. Ms. Alexander explained, "I ask, 'Why do you think you did it well? How did you get through that question? What skills did you use?' And with the kids who struggle, we talk about what the question means and discuss reasons they might have gotten stuck" (T99F). In a later interview she noted, "When we first started doing those samples, they would say things like, 'I can't even read this' or 'This is way over my head.' And I said, 'Exactly. And that's going to happen on the test'" (T00S). The NCS Mentor program is another valuable resource in Ms. Alexander's efforts to help students prepare for the WASL. She explained, "I put it up on the TV screen so they can all see it. Then we read the papers (on the screen) and the students help me score them. We go back to the NCS

Mentor to see how the state scored those papers, and we see how our scores matched. Then, I try to tie it back in to their writing” (T99F).

Like Ms. Alexander, Ms. Watson is deliberate and systematic in helping students prepare for the mathematics component of the WASL. She makes extensive use of the state’s scoring rubric. “I have conversations with them about what the scoring on the WASL is going to look like. Early in the year we talk about what it looks like to score at the level of a one or a two. And then we talk about what to do to raise that score. If we use threes and fours as examples, we get together in pairs or as a whole class and talk about their justification for that level” (T00S). Later in the interview, she commented, “The students are really making some positive changes. You can just see how motivated they are, and how they’re so much more able to communicate” (T00S).

In contrast to patterns revealed on the surveys, however, we did not notice an increase in test preparation activities at our exemplary schools as the time of the WASL drew near. On the contrary, successful student performance seemed to motivate teachers at these schools throughout the year. Thus, Ms. Erikson, Ms. Alexander, and Ms. Wright all organized their writing instructional programs around the genres identified in the EALRs. And, as we noted above, Ms. Wright focused almost exclusively on narrative and expository writing from the beginning of the school year until the time of the WASL. Similarly, throughout the year Ms. Erikson, Ms. Watson, and Ms. Wright had their students write about mathematics on a daily basis.

Each school also provided schoolwide programs geared toward WASL preparation. In the fall of 1999 Emerald instituted “WASL Fridays” for all fourth graders. Ms. Glen explained, “In the fourth grade every Friday is WASL preparation. They look at really specific skills and help the students prepare for the WASL. That started in September” (P99S). Vista established an extended learning program in response to the district’s accountability plan that passed in the spring of 1999. Ms. Thompson elaborated, “All of our children who are 2 or more years below grade level or who demonstrated poor performance on the WASL are asked to be in an extended learning program of some sort. They either have to come in the mornings from 8:00 to 8:30, go during their lunchtime from 1:00 to 1:30, or stay after school from 3:30 to 4:15. They have to be involved in some program” (P99F). This attention to the WASL throughout the year may be one of the characteristics that set apart exemplary schools in their reform efforts.

Stecher and Chun (2001; see also Stecher et al., 2000) raised a concern about the extent to which attention to the WASL rather than the EARLs may narrow the instructional focus at Washington schools from the broad set of domains encompassed by the standards to the more limited set of test specifications and item formats covered by the test. As our discussion of the two themes related to assessment indicates, all three case study schools took the WASL into account in making curricular and instructional decisions. They provided practice with the specific item formats students would encounter on the test and, in some instances (e.g., Ms. Wright’s focus on the two tested genres), they limited the curriculum based on WASL content. However, their decisions based on analyses of WASL scores concerned broad areas of instructional emphasis in listening, reading, writing, and mathematics. For the most part, “teaching to the test” at these exemplary schools did not represent a narrowing of the curriculum or time devoted to test preparation at the expense of broader learning goals—outcomes that have been identified as potential undesirable consequences of high-stakes testing (Stecher & Barron, 1999; Stecher & Mitchell, 1995). Thus, the case study data provide a more nuanced interpretation of the survey finding that the WASL had a greater impact on schools than the EARLs—an interpretation that suggests a more complex relationship among standards, high-stakes assessments, and curriculum and instruction than either the cases or the surveys alone could provide.

Discussion

These results illustrate some of the advantages of combining case study and survey techniques in studying school reform. We were able to better describe the impact of the Washington reform on school and classroom practices when we drew on both sources of data. We could begin to unpack explanations for survey findings using understandings gleaned from the cases. And we identified ways in which exemplary schools were typical of reform efforts across the state and ways in which they were different.

We began with simple descriptive links between survey findings and case studies, instances in which the case studies provided concrete examples of actions described in the aggregate in the surveys. These examples, such as descriptions of specific curriculum alignment activities and ways of integrating the teaching of writing conventions, the writing process, and genre, provided useful information about *what* schools and teachers were doing to address the goals of the Washington

reform. As a result, we were better able to interpret the summary data from the survey in the context of these exemplary schools. Such elaborations of *what* occurred are a common way in which surveys and case studies complement each other.

In some instances, the case studies also enabled us to make assertions about *why* teachers acted as they did or *how* their environment influenced their actions. Such explanatory elaborations included, for example, the reasons for Emerald’s initial focus on reforming mathematics instruction (because of a district grant for mathematics professional development) and Vista’s emphasis on language arts (because of the principal’s beliefs about the central role of literacy in school success). Of course, we were not able to generalize these explanations to the state as a whole; we had only a few selected (and nonrepresentative) cases to draw upon. Nonetheless, the ability to draw on multiple sources of evidence to develop understandings of specific instances of actions whose statewide prevalence we knew helped us to generate possible interpretations of patterns in survey results—in this case, the variety in instructional content and practices.

Finally, some of the practices we observed in the case study schools could be situated in the larger state context by comparing their prevalence or intensity to the distribution of survey responses. These comparisons provided insights about what makes the case study schools exemplary. As one example, case study schools did not increase test preparation activities as the time of the WASL drew near; rather, classroom-level practices such as the incorporation of tool kit activities in mathematics and schoolwide programs such as “WASL Fridays” were regular features of the schools’ instructional programs throughout the year.

By cooperating in design and analysis, our case study and survey teams were able to develop better understandings of the impact of standards-based reform at the school and classroom levels. Thus, our study of the Washington education reform serves as a “proof of concept” that these methods can be used in complementary ways that are more powerful than either approach used alone.

However, we are equally certain that we did not achieve the full potential of this combined strategy. There were a number of survey findings for which the case studies offered no explanation; for example, what was occurring in schools that took little or no action to improve student performance? Similarly, several case study insights could not be put in larger context using survey results. For example, how widespread were systematic efforts to use test results for curriculum planning and

instructional improvement? In addition, there were some instances in which the case studies illuminated *what* occurred in greater depth and with more clarity than the surveys alone, but did not help us to understand *why* relationships existed or *how* one factor influenced another. For example, the surveys provided some evidence that attention to genre was a relatively recent change in writing instruction but attention to the writing process was longstanding. We looked to the case studies for clarification or confirmation but found little. This was not an issue we explicitly addressed in the case analyses and write-ups. Such limitations led us to contemplate ways we could increase the utility of such joint efforts in the future. Smith (1997) offers a clear example of how assertions about events can be validated using multiple data sources. For example, her database was rich enough to provide both convergent and divergent evidence to establish the warrant for her assertions regarding the five different ways in which educators “understood ASAP.”

There were some instances in which survey and case study data supported possibly contradictory interpretations. For example, we collected much evidence that teachers were refocusing their instruction in response to the Washington education reform. One possible interpretation is that they were attending to the content of the WASL and ignoring aspect of the EALRs that were not tested. Another possible interpretation is that teachers were working to promote mastery of the EALRs and were using the WASL to help them identify areas of weakness in students’ skills and understandings. Data from the surveys are consistent with the former view; evidence from the cases supports the latter interpretation. Unfortunately, we cannot resolve this apparent contradiction. In fact, this demonstrates how the choice of method can influence the results of the research. The survey questions were framed as selected-response items, so teachers had to choose among a small number of fixed alternatives written by the researchers. In the case studies, teachers expressed their views in their own words, and they provided explanations that did not necessarily fit neatly into the options provided in the surveys.

In thinking about how we might have made our efforts more effective, we identified some structural and some conceptual features that could be modified to improve the “yield” of studies employing multiple methods. The first of these is particularly relevant to our situation, in which responsibility for the survey and case study components of the project was divided between researchers at two different institutions. Neither Smith (1997), Spillane and Zeuli (1999), Borman and Lee (2001),

nor Schorr and Firestone (2001) had to deal with this additional complication. In addition to whatever stylistic or epistemological distance may separate survey and case study researchers, the physical distance that separated the teams in our study further hindered insightful and analytically rich communication. We used electronic mail and the telephone for regular exchanges, and we met together for two days at the beginning of the Washington project to coordinate designs. Yet, these interactions tended to focus on the logistical features of the study—schedules, monitoring reports for CRESST, communication with state officials, coordination of visits to sites, and so on. What was lacking, in our opinion, were rich conversations about emerging results, unanswered questions, and evolving insights. This sort of exchange often occurs informally in the course of working with data, and it occurred among the members of the survey team and the case study team, respectively. It occurred much less often between the two teams.

We do not have an easy solution to this limitation; ideally we would recommend adapting the physical and temporal arrangements so the researchers interact directly on a regular basis as they review and understand the information they have collected. In practice, however, such arrangements are not always possible. When distances intervene, researchers must make a conscious effort to devote more time to “high-bandwidth” interchanges. New technologies are being developed that may foster such collaboration. RAND has installed videoconference facilities in all its offices across the country, and researchers use these regularly for “face-to-face” meetings. Unfortunately, this equipment is expensive, all institutions do not provide it, systems are not all compatible, and the operating cost can be high. Nevertheless, it has advantages over electronic mail and telephone interchanges.

A second way in which we could have improved our study would have been to develop a design more sensitive to the demands of different data collection methods and that better facilitated sequential refinement. Each of our two teams refined its focus and redirected its efforts in the second year of the study based on findings from the first year. However, there was little feedback between the teams to make these second-year plans responsive to the totality of the first-year results. One reason tighter integration failed to occur was the natural timelines associated with the two data collection strategies. Surveys can be collected, coded, and analyzed more quickly than case studies. In our situation, preliminary results from the first-year surveys were available at the time the second-year surveys had to be developed, but preliminary results from the case studies were not. To be responsive

to the school calendar, we adopted a tight, annual survey cycle. We designed the surveys in the fall and winter, collected the data in the spring, conducted initial analyses during the summer, and had findings to feed back into the next round of design the following fall. The case studies had a much different rhythm. We visited each school in spring 1999, fall 1999, and spring 2000. Between visits we transcribed interviews and summarized field notes. Each researcher conducted a preliminary analysis of data from her site to inform subsequent site visits. However, we did not begin the kind of cross-case analysis that could have informed survey development. As a result, although we were able to use preliminary results to refine data collection within each method, it was difficult to utilize insights gained from one method to inform the ongoing refinement of the other.

To overcome this problem, we recommend adopting an initial research plan that is sensitive to the inherent rhythms of the two approaches. Rather than conducting surveys annually, it might make more sense to schedule surveys in the first and third years of a multi-year project. This would allow the design of the second survey to be informed by the results of the case studies. Similarly, rather than beginning all case studies in the first year and continuing them into the second, it might make sense to initiate half of the case studies in the second year. This would allow selection of some sites to be informed by patterns revealed in the surveys.

Even when timing would have allowed for more cross-fertilization, we did not always avail ourselves of the opportunities to refocus either research strategy, because our design called for maintaining the same data collection strategy over time. One of the tradeoffs difficult to negotiate was between consistency of data collection (which is important for portraying changes over time) and responsiveness to new information (which is important for illuminating emerging findings). We typically held consistency as a higher goal than responsiveness. Although there is no easy solution to this dilemma, we would recommend being more receptive to modifying data collection to take greater advantage of multiple methods, even at the expense of potentially sacrificing some ability to portray changes over time.

If resources permitted, we would have preferred a more diverse set of case study sites. Given resource constraints, we opted to focus the case studies on exemplary sites. We made this choice because we thought the most important thing to understand was how effective schools were responding to the pressures of the reform. However, the surveys made it clear that after two years there was wide variation in implementation, and many schools were not making equal progress.

Under these circumstances, and given the complexity of systemic reform, understanding the constraints on schools that exhibit uneven patterns of implementation becomes a relatively more important issue. As Cronbach and Associates (1980) note, “If nothing else, a closer look at the less successful realizations can suggest guidelines that will make such deviations infrequent” (p. 278). We would recommend using survey data to inform site selection in ways that produce more a more diverse sample.

Finally, our work points to the need for better strategies for collaborative analyses. The image of complementary analyses we described in our original research proposal proved to be difficult to achieve in practice. To a certain extent we did use the surveys to “portray the landscape” and the case studies to “illuminate key locations,” but only occasionally did we link these two images in conceptually rich ways. We attempted to have the analyses support one another in a couple of ways. First, we agreed to address common themes when we designed the study. Second, we met midway through the investigation to share results from our separate analyses and identify areas of overlap and points of disagreement. Each team summarized its findings so that we could look for areas of support. Then each raised questions that the other might have information to help answer. These *ad hoc* strategies were helpful, but limited. They pointed out that despite our initial intentions to examine similar questions, we did not obtain as much complementary data as we had hoped. More attention to coordinating our efforts, both as we designed our data collection systems and developed our analysis strategies, would have been beneficial.

Conclusions

Our research shows that multiple research methods are beneficial when studying complex processes like school reform. It also points out that successful integration of case study and survey methods must be cultivated at the design stage when conceptual frameworks are developed and data collection strategies are planned, at the sampling stage when sites are selected, and at the analysis stage when multiple sources of data are integrated to explore assertions about thoughts, actions, and relationships. In our case, the need for extensive, ongoing communication among researchers was perhaps as important as any of these structural improvements.

References

- Borko, H., Wolf, S. A., Simone, G., & Uchiyama, K. (2001, April). *Schools in transition: Reform efforts in exemplary schools of Washington*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Borko, H., Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two teachers' experiences in an intensive mathematics and literacy staff development project. *Elementary School Journal, 100*, 273-306.
- Borko, H., & Elliott, R. (1999). Hands-on pedagogy versus hands-off accountability: Tensions between competing commitments for exemplary math teachers in Kentucky. *Phi Delta Kappan, 80*, 394-400.
- Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13*, 259-278.
- Borman, K. M., & Lee, R. (2001). *Linkages among professional development, classroom practice, and student outcomes*. Unpublished manuscript, University of South Florida, Tampa, FL.
- Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Fullan, M. G., & Miles, M. B. (1992). Getting reform right: What works and what doesn't. *Phi Delta Kappan, 73*, 745-752.
- Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.
- Knapp, M. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 67*, 227-266.
- Koretz, D., Stecher, B. M., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practices, 13*(3), 5-16. (Reprinted as RP-366, 1995, Santa Monica, CA: RAND)
- Linn, R. L. (1990). *Quantitative methods*. Washington, DC: American Educational Research Association.
- McCourt, B., Boydston, T., Borman, K., Kersaint, G., & Lee, R. (2001). *Instructional practices of teachers in four urban systemic initiative sites*. Unpublished manuscript, University of South Florida, Tampa, FL.

- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teaching? *Educational Researcher*, 21(1), 4-15.
- Schorr, R. Y., & Firestone, W. (2001, April). *Changing mathematics teaching in response to a state testing program: A fine-grained analysis*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Shulman, L. S. (1983). Autonomy and obligation: The remote control of teaching. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 484-504). New York: Longman.
- Smith M. L. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program (ASAP): Equity and teacher capacity building* (CSE Tech. Rep. No. 425). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21, 1-28.
- Stecher, B. M. (1998). The local benefits and burdens of large-scale portfolio assessment. *Assessment in Education*, 5, 335-351.
- Stecher, B. M., & Barron, S. (1999). *Quadrennial milestone accountability testing in Kentucky* (CSE Tech. Rep. No. 505). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- Stecher, B. M., Barron, S. L., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms* (CSE Tech. Rep. No. 525). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Tech. Rep. No. 482). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- Stecher, B. M., & Chun, T. (2001, April). *The effects of the Washington education reform on school and classroom practice, 1999-2000*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio driven reform; Vermont teachers' understanding of mathematical problem solving* (CSE Tech. Rep. No. 400). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Wolf, S. A., Borko, H., Elliott, R., & McIver, M. (2000). "That dog won't hunt!": Exemplary school change efforts within the Kentucky reform. *American Educational Research Journal*, 37, 349-393.

Wolf, S. A., & McIver, M. (1999). When process becomes policy: The paradox of Kentucky state reform for exemplary teachers of writing. *Phi Delta Kappan*, 80, 401-406.