CRESST REPORT 742

Joan L. Herman
Kyo Yamashiro
Sloane Lefkowitz
Lee Ann Trusela

EXPLORING DATA USE AND SCHOOL PERFORMANCE IN AN URBAN PUBLIC SCHOOL DISTRICT

SEPTEMBER, 2008

National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

Exploring Data Use and School Performance in an Urban Public School District

Evaluation of Seattle Public Schools'
Comprehensive Value-Added Assessment System

CRESST Report 742

Joan L. Herman, Kyo Yamashiro, Sloane Lefkowitz, and Lee Ann Trusela
CRESST/University of California, Los Angeles

September 2008

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work and research reported herein was commissioned and supported by the Stuart Foundation, Award Number 0008258-001.

The findings and opinions expressed herein are those of the authors and do not necessarily reflect the positions or policies of the Stuart Foundation.

EXPLORING DATA USE AND SCHOOL PERFORMANCE

IN AN URBAN PUBLIC SCHOOL DISTRICT[1]

Evaluation of Seattle Public Schools'
Comprehensive Value-Added Assessment System

Joan L. Herman, Kyo Yamashiro, Sloane Lefkowitz, and Lee Ann Trusela
CRESST/University of California, Los Angeles

Abstract

This study examined the relationship between data use and achievement at 13 urban Title I schools. Using multiple methods, including test scores, district surveys, school transformation plans, and four case study site visits, the researchers found wide variation in the use of data to inform instruction and planning. In some cases, schools were overwhelmed with the amount of data or were not convinced that alternating test score data from two different tests provided dependable information. The researchers did not find a substantial link between data use and achievement, which may have been a result of the small sample size or different implementation methods between schools. Teachers and principals recommended important needs for more timely data delivery, individual versus group data reports, and better training in assessment and data analysis.

INTRODUCTION

Data use and evidence-based practices are at the heart of current school reform efforts. Fueled by the Federal No Child Left Behind Act (NCLB, 2002) and state and district policies, schools and the administrators and teachers within them are expected to use assessment data to identify student needs, to formulate school goals, to plan and implement educational strategies to achieve those goals, to monitor progress, and to continue to revise and refine their efforts to improve the academic performance of the school as a whole and each of the subgroups and individual students that comprise it. The logic is appealing: schools will use data to engage in a continuous improvement process, and indeed the cycle is one that has been long advocated for education (see for example, Tyler, 1949) and more recently in business (Deming, 1982), where the power and success of "learning organizations" that infuse data throughout their decision making processes has been widely recognized (Senge, 1990).

[1] Special thanks to Rachel Montgomery and Kristine Chong for editing and formatting support.

Yet despite the logical appeal, the evidence on the extent and effects of data use is weak. Available research, in fact, suggests that some of the prerequisites to effectively integrating data in schools' and teachers' decision-making processes may be problematic. For example, earlier studies have shown the limits in schools' access and capacity to analyze available data in leadership and cultural support for change (Herman & Gribbons, 2001). Research on teachers' use of assessment reveals similar issues in teachers' use of classroom data to inform teaching and learning: Teachers have not been trained in assessment and many lack the sophisticated content and pedagogical knowledge needed to interpret student performance and pursue effective instructional alternatives (Herman, Osmundson, Ayala, Schneider, & Timms, 2005; Heritage & Yeagley, 2005).

The study reported below sought additional information about the role of data in school improvement and the factors and strategies that support their effective use. We started with the supposition that if data use is critical to school improvement, then data use practices ought to differentiate effective from less effective schools and set out to test that hypotheses. In collaboration with a large urban district in the Pacific Northwest, we identified elementary schools serving low socioeconomic status (SES) students that were "beating the odds" in terms of fostering growth in their students' academic performance and a comparison group of schools whose growth patterns were more typical for the district. Our earlier study had revealed that schools with higher concentrations of low SES students showed relatively less growth in student achievement relative to the district average, and that low performing students in these schools showed relatively less growth than their initially middle- and high-performing peers (Choi, Seltzer, Herman & Yamashiro, 2004). The current study used multiple methods to examine the data use practices within these *Beat the Odds* and typical schools in an attempt to explore the practices and factors that might contribute to school success. In the sections that follow, we describe our methodology and data sources, present results derived from them, and conclude with implications and next steps for research and practice in data use.

METHODOLOGY

Sampling was based on longitudinal student data available from the district. Study data sources and procedures included reviews of school transformation plans; observations of school presentations about their progress; special interviews and surveys conducted in the course of site visits; and available district survey data related to issues of school climate and culture.

Sampling

Sampling was based on available student-level, longitudinal data for the Iowa Test of Basic Skills (ITBS) in reading and math for the years 1998 to 2003. Latent variable, multilevel analyses were used to estimate average gains for each school in the district as students progressed from 3rd grade to 5th grade, for two cohorts of students—those who were 3rd graders in 1998 and those who were 3rd graders in 2001. The analyses explored school growth trajectories for students at three different levels of initial achievement:

- Average: students who started at the mean of their schools' performance on the ITBS
- Low: students who started at 15 points below the school mean; and
- High: students who started at 15 points above the school mean.

Based on these analyses, we identified Beat the Odds schools (n = 7) who were below average in SES, who showed higher than average growth trajectories for the school as a whole and or for students who were low in initial achievement status and who were relatively consistent in performance in both reading and math and for the two cohorts. We identified as comparison schools six schools that were demographically similar to the Beat the Odds schools (e.g., in percentage of free lunch, ethnicity) and in terms of initial achievement status.

All schools in the sample were Title I schools and were ethnically diverse, more so than the district as a whole. As seen in Table 1, the percentage of White students across these 13 schools ranged from 3% to 59%. With most of the 13 schools (11), White students were a small minority, less than 25% of the school. Five schools had African American student populations of between 25% and 53%, whereas three schools (Pierce, Truman, and Polk) had African American student populations between 76% and 81%. Similarly, there were significant Asian student populations in seven of the schools, ranging between 25% and 57%. While Latino populations generally were similar to the district average, more than half the schools in our sample (7) had bilingual populations of more than double the district average (25% to 44%).

Table 1

2004 Demographics: Student Ethnicity

School	Total	Grade level	American Indian	Asian	African American	Latino	White
District	46,416	K–12	2	23	22	11	41
Van Buren	299	Pre K–5	2	17	13	9	59
Carter	370	K–5	1	54	11	25	10
Harding	258	K–5	2	40	40	15	3
Hoover	533	Pre K–5	3	21	16	16	44
Fillmore	169	K–5	1	18	33	21	26
Jefferson	167	K–5	4	38	34	19	5
Kennedy	430	K–5	4	34	15	29	17
Lincoln	519	K–5	1	59	11	8	22
Pierce	232	K–5	1	8	76	9	6
Polk	134	K–5	1	2	81	10	5
Truman	208	K–5	2	2	80	7	9
Tyler	253	K–5	2	26	53	12	6
Wilson	295	K–5	2	58	25	10	5

Note. With the exception of "Total," and grade level, all other numbers are percentages. K = kindergarten.

In terms of special populations served, these schools tended to be more diverse than average as well. Schools in this sample had wide variation in their bilingual populations (from 0% to 41%). The district average bilingual population was 12%. More than half the schools in our sample (7) had bilingual populations of more than double the district average (25% to 44%). In addition, approximately 9% of district students are special education students and three of the sample schools served populations just slightly larger than the district average (10% and 11%): Fillmore, Hoover, and Tyler.

Transformation Plan Rating Process

Because we thought school improvement planning offered a window into schools' use of data in decision making, we conducted reviews of the transformation plans that every school in the district was required to submit. The review encompassed 3 years of plans for our sampled schools, those submitted in anticipation of the 2002–2003, 2003–2004, and 2004–2005 school years.

Transformation Plan Rubric

Based on a literature review of data use in schools and school performance indicators, a rubric was developed to address five primary components of data use:

1. Types of evidence or indicators used;
2. Identification of goals/objectives through needs analysis;
3. Identification of solution strategies;
4. Analysis of progress; and
5. Inclusion of stakeholders.

Within each of these main components, the rubric provided for ratings on one to four specific quality dimensions:

Component 1: Types of evidence or indicators used. The four dimensions captured by this component included: (a) breadth/range of evidence, (b) depth of analysis, (c) use of value-added data, and (d) technical sophistication.

The breadth rating considered the number of different sources of information used by the school, ranging from reliance on state test results only to a set of evidence, such as parent survey data, classroom observations, classroom-based assessments, and portfolios. Depth addressed the detail at which schools analyzed each of their sources (e.g., ranging from just noting the level of particular scores to examining subject matter performance in relation to trends, in comparison to other subjects, and by subscales). Use of value-added data captured whether schools mentioned the value-added data in their planning and their perceived understanding of it. Finally, the technical sophistication rating addressed the appropriateness of the school's data analysis strategies.

Component 2: Identification of goals or objectives through needs analysis. This component was intended to measure the link between the school data and the types of goals and objectives set out in the transformation plan. The rating addressed the extent to which school goals were rooted in data on student needs and in logical remedies for those as opposed to goals and objectives seeming a "hodgepodge," without much rhyme or reason.

Component 3: Identification of solution strategies. The two dimensions that measured component three are: (a) specificity and (b) theory- or research-based/data-driven.

The specificity dimension referred to the concreteness with which the school articulated its solution strategies. We looked for the extent to which general strategies were accompanied by action plans and specific benchmarks. The theory-based dimension referred

to whether the school identified strategies ad hoc, or identified strategies based on some theory of change, on a review of the literature, or on available evidence of effectiveness.

Component 4: Analysis of progress. This addressed the degree to which schools planned for formative or periodic assessments by which they could monitor their process. In this component, we were looking for whether the school planned to periodically review data or other evidence of progress to make mid-course corrections, if need be. Essentially, we looked for evidence that the school planned to review data more than once a year and in addition to the annual state test results.

Component 5: Inclusion of stakeholders. This component encompassed the degree to which various stakeholders were included throughout the transformation process. In other words, we looked for evidence that stakeholders (e.g., parents, Building Leadership Team [BLT] members, grade-level chairs, teacher aides, and community members) were part of the planning process.

Improvement. The rubric thus addressed nine dimensions over the five components. Because each of these dimensions was rated for the three plans, we also could examine whether or not schools' data use, as evidenced in the plans, was changing or improving over the period. To do so, we computed difference scores, comparing the first 2 years (because the 2 were integrally related) to the third year.

Transformation Plan Review Process

Three researchers were trained to use the rubric to rate the transformation plans. An inter-rater reliability exercise was conducted on one randomly selected school to calibrate ratings. For this exercise, after initial training, raters independently reviewed the same school's transformation plan. Once complete, the three raters met to discuss and resolve their individual ratings. Raters were asked to provide a rationale for the scores they chose, particularly when there were discrepancies between raters. This consensus-building process helped to assure that raters were each using consistent criteria. From this point, the plan sample was divided up between researchers so that two raters reviewed each of the remaining plans. After the ratings were complete, any discrepancies were resolved by subsequent discussion and consensus.

District Survey Analysis

Additional information was available for all 13 schools from the district's annual survey of staff climate. The survey queried respondents about school leadership, instructional planning and teacher collaboration, feelings of trust and respect, communication, and

included a few questions related to data use. Items were grouped and aggregated according to these themes. Because we had only school-level averages from the web site reports, we were not able to empirically validate these subscales (e.g., through factor analysis or some other data reduction technique) that were so created. Two researchers individually grouped the items based on substantive content and then resolved any differences through consensus.

Once items were grouped according to themes, we constructed averages across those items, based on the 5-point Likert scales used by the district. Most of the items used one of two variants of a 5-point scale—reflecting either strength of agreement with a particular statement or opinion from 1 (*strongly disagree*) to 5 (*strongly agree*); or behavior frequency 1 (*never*), 2 (*a few times a year*), 3 (*once or twice a month*), 4 (*once or twice a week*), or 5 (*almost daily*). One set of questions (items 4a–4d) used a 5-point scale representing the degree to which teachers felt they had influence over particular decisions: 1 (*no influence*) to 5 (*a great deal of influence*). Because of the similarity of the scale ranges and in their direction 1 (*low, negative*) to 5 (*high, positive*), and in the absence of other reliable alternatives, we combined items across these different response scales. Moreover, although we recognize the limits of constructing averages from scales that are not equal interval or that do not represent a continuous variable, it was the best analysis that could be conducted based on available data—that is, only school level averages.

A second major limitation of the survey data was the comparability of the responses from year to year. Although the same questions were asked, the response rates each year could not be established, and it is very likely that the respondents varied from one year to the next. As a result, it was impossible to determine whether changes in responses over time were the result of changes in school climate or practice or simply were the result of changes in respondents. The inferences that can be drawn from these data thus are very limited, and we used them only to identify major differences among the sample schools. For the most part, the responses of the 13 schools in our sample tended to follow district averages.

Case Study Site Visits

We planned to conduct site visits to all 13 schools in our sample. However, because of changes in district leadership and its commitment to the study and because the timing of our visits overlapped with the district testing window, only four sites agreed to participate in this portion of the study. We visited each of these four schools for a 2-day period in May or June 2004 during which specially developed interviews and surveys were conducted. Interviews were scheduled with the principal, the school's BLT, and two grade-level groups of teachers divided by grade levels (kindergarten to 2nd grade and 3rd grade to 5th grade). Typically two

researchers conducted the interviews and tape-recorded the conversation for future transcription. All four interviews took place at three of the sites (Van Buren, Wilson, and Polk). Only three interviews took place at Truman because the school's BLT is comprised only of teachers, all of whom were included in the teacher interviews. Each principal was interviewed individually; BLT interviews included 6–9 participants; and the group teacher interviews varied from 2–6 teachers at a time. In addition to the interviews, surveys were distributed to all teaching staff prior to the visit. To ensure anonymity, a return envelope was included, but most teachers opted to hand in the surveys to the researchers on site. Response rates at the four sites were: 39% at Van Buren; 36% at Wilson; 26% at Polk; and 56% at Truman (see Table 2).

Incentives were provided for survey completion (a $5 gift certificate to Starbucks), to improve response rates. As another token of appreciation for participation, we also provided all of the staff who participated in an on-site interview with a $50.00 check. School sites received an honorarium of $500.00 for their participation in the study.

Table 2

Case Study Survey Response Rate By School

School Site	Certificated staff	Classified staff	Total staff	No. surveys returned	Response rate
Van Buren (2)	22	9	31	12	39%
Wilson (17)	19	9	28	10	36%
Polk (23)	14	17	31	8	26%
Truman (40)	21	13	34	19	56%
Total	76	48	124	49	40%

All interviews were transcribed and qualitative analysis software was used to code the data. To analyze the interviews, researchers created codes and sub-codes using an inductive approach. A general review of the transcriptions informed the development of a code set that reflected salient concepts and common responses across interviews and respondents. The Atlas.ti statistical software package was used to code the interviews. Coding reliability was attained through researcher consensus. Two researchers individually coded each of the interview transcripts. After coding was complete, researchers shared their ratings, and final reliability was attained through consensus.

Simple frequencies and cross-tabulations were computed for the teacher survey data. Survey data were aggregated at the school site as well as across the four sites, to determine overall trends. School site aggregates were compared to overall averages, to gauge whether any particular sites varied significantly from the overall average. Summaries of the data from each of the four case study sites are found in the appendix.

In the Results section of this report, we indicate the data sources from which we were drawing, since we have more extensive case study data for only 4 of the 13 schools. The codes of the data sources referenced are shown in the Table 3.

Table 3

Data Source Codes

Data source	Code
CRESST case study interviews	CS interviews
CRESST case study survey	CS survey
District staff survey	District survey
Transformation plans	TP
Transformation plan ratings	TP ratings

Achievement Analyses

As mentioned above, the 13 schools in this study were initially selected based on their performance across two time points. Specifically, the sample contrasted low SES schools that had made larger than average gains in the performance of students from 2nd grade to 5th grade on the ITBS for the period just prior to the study with those that had shown more typical growth trajectories for the district. As the study was conducted, we sought to confirm the stability of our sample designation with additional data, and examined performance over five different cohorts for each school, from those who started in 3rd grade in 1998 to those who were in 3rd grade in 2001. These multiple cohorts also allowed us to examine a longer-term picture of student performance in each school:

- 1998: Growth of students from 3rd grade in 1998 to 5th grade in 2000;
- 1999: growth of students from 3rd grade in 1999 to 5th grade in 2001;
- 2000: growth of students from 3rd grade in 2000 to 5th grade in 2002;
- 2001: growth of students from 3rd grade in 2001 to 5th grade in 2003; and
- 2002: growth of students from 3rd grade in 2002 to 5th grade in 2004.

For these analyses, we conducted latent variable, multilevel analyses to estimate the average gains experienced by each cohort of students in each school in reading and math on the ITBS, using student-level, longitudinal data across 7 years (1998–2004). We call this design the Latent Variable Multiple Cohort (LMC) Analysis.[2] We estimated average (or value-added) gains across two time points (Grades 3–5) on the ITBS for each successive cohort. We recognize that the ITBS is arguably less sensitive to instruction than the Washington Assessment of Student Learning (WASL), Washington State's standards-based assessment. We used ITBS scores, however, because ITBS is administered in two elementary grades: 3rd grade and 5th grade; whereas, WASL is only administered in the 4th grade, so gains could not be estimated on the WASL for elementary schools. In addition, we used the ITBS because we were able to retrieve standard errors of measurement for scores attained in both grade levels. This allowed us to account for measurement error in the test, which provides more precise estimates of schools' gains.

Moreover, in addition to examining overall school gains, we also examined expected gains by different levels of initial student achievement status to helps us understand how progress varies by relative achievement level. The analyses thus address growth as well as the equitable distribution of growth across subgroups, which are defined by their initial performance status rather than by their ethnic, language, disability, or income levels. For example, we completed expected growth patterns for students who performed at a school's mean initial starting value, as well as for students who performed 15 points above and below the school's average starting point. As described in the Results section of this report, these subsequent analyses revealed significant problems in the stability of school designations as Beat the Odds.

[2]We use an advanced 2-level hierarchical modeling technique in a fully Bayesian framework (Seltzer, Choi & Thum, 2003; Choi & Seltzer, 2003; see also Raudenbush & Bryk, Chapter 11, 2002 for a maximum likelihood based approach). This modeling methodology incorporates latent variable regression techniques into 2-level hierarchical models. In a Level-1 (within-individual) model, time-series observations are modeled as a function of a time metric, and we obtain estimates of initial status and rates of change for individual students. In a Level 2 (between-individual) model, rates of change for each student are modeled as a function of their initial status. The coefficient capturing the relationship between initial status and rates of change represents an expected increase (in the case of a positive coefficient) or decrease (in the case of a negative coefficient) in rates of change when initial status increases one unit. We apply this modeling technique to settings where we have observations for individuals at two points in time, and the corresponding standard errors of measurement for both observations.

RESULTS

Our data suggest that sampled schools claim to use a variety of data for a variety of purposes, including planning and instructional decision making at the school, classroom, and individual levels, although their explanations of data use practices tended to be fairly cursory. Analyses showed more similarities than differences in the general use of data across the 13 schools, as described below.

Note that among the four case study sites, where the data are richest, each school has a unique story about the ways in which data are used to inform planning and to monitor and improve student progress. Some of this information is used to illustrate the themes found across all 13 sites. For a more in-depth description, however, see Appendix C for individual vignettes.

Data Use in School Planning

Our review of Transformation Plans provided basic data on whether and how sampled school incorporated data during school planning over the course of the study. Because we summarized results at two time periods, our analyses also addresses whether and how data use might be changing.

Overall Trends

Table 4 summarizes results across our 13 schools by year on the various components of our plan review. The results show that schools' use of a variety of indicators to support their school planning started and continued as a relative strength in their use of data in school planning. Content analysis of the plans, interviews, and surveys during site visits suggested that all schools referred to the use of state standardized test results, including WASL and ITBS. Most schools also cited the use of district assessments such as the Developmental Reading Assessment (DRA) and Direct Writing Assessment (DWA) and by the second year, most made mention of curriculum based testing—or curriculum-based assessment (CBAs). Indeed during site visits, teachers mentioned that these latter sources—DRA, SWA, and CBA—were of most use in identifying student and curriculum needs (TP, CS Interviews, CS Survey, and TP Ratings).

Whereas the use of value-added data was a relative weakness in the first-year plans, this was an area in which schools showed improvement. By Year 2, schools at least mentioned the value-added results as part of their planning. Schools also showed some improvement in the specificity of their action planning. In contrast, use of data in analyzing progress and inclusiveness in decision-making remained stable or decreased across the

period. Although this decrease in inclusiveness was very small, interview data from the four case study sites echoed the trend. Interview data indicated that at two school sites (Wilson and Van Buren) the task of school planning had became more centralized over time, with broad participation less necessary or desired.

Of particular interest in Table 3 are those dimensions for which the large majority of schools scored at the mode. For example, for the second-year plans, 85% of schools clustered at the mode for breadth and coverage of indicators (3.00), use of value-added data (2.00), and inclusiveness of stakeholder groups (2.00). These results indicated that the Year 2 plans (2003–2004), showed that most schools used a breadth of indicators and at least mentioned their value-added data (though most did not give very detailed analysis of this data); and that most schools had at least minimally involved various stakeholders, with some mention of the BLT, Parent Teacher Association (PTA), and other common groups, but did not systematically involve teacher groups or grade-level teams in decision making.

Table 4.

Mean, Mode, and Percentage at Mode by Dimension-Transformation Plan Review

Quality dimension	Year #	Mean	Mode	% at Mode
Types of indicators	1	1.81	1.75	30.8%
	2	2.33**	2.25	46.2%
Breadth/coverage	1	2.31	2.00	53.8%
	2	2.85*	3.00	84.6%
Depth	1	1.38	1.00	61.5%
	2	2.23**	2.00	46.2%
Value-added	1	1.62	2.00 [a]	46.2%
	2	2.00	2.00	84.6%
Technical sophistication	1	1.92	2.00	92.3%
	2	2.23	2.00	76.9%
Objectives linked to needs analysis: *Breadth/depth*	1	2.00	2.00	53.8%
	2	2.31	2.00	69.2%
Solution strategies	1	1.96	2.00	69.2%
	2	2.19	2.00 [b]	38.5%
Specificity	1	2.00	2.00	69.2%
	2	2.38	2.00	61.5%
Theory/research/data driven	1	1.92	2.00	92.3%
	2	2.00	2.00	69.2%
Analysis of progress: *Formative/periodic*	1	2.08	2.00	61.5%
	2	2.31	2.00	69.2%
Inclusiveness: Throughout process	1	2.08	2.00	76.9%
	2	2.15	2.00	84.6%
Total	1	1.74	1.79 [c]	15.4%
	2	2.00	2.00 [d]	15.4%

Note. # = Denotes the academic year of the transformation plan (Year 1 = 2002–2003 & 2003–2004, Year 2 = 2004–2005).
[a] Multiple modes: Value listed is closest to the mean. Modes: 1.00, 2.00.
[b] Multiple modes: Value listed is closest to the mean. Modes: 2.00, 2.50.
[c] Multiple modes: Value listed is closest to the mean. Modes: 1.79, 1.86, 1.93.
[d] Multiple modes: Value listed is closest to the mean. Modes: 1.64, 1.79, 2.00.
* $p < .01$

The increased attention to value-added data is worth underscoring, because of its special interest to the original project goals. Table 5 shows almost all schools made at least *some* use of value-added data in the second year of analysis, a substantial increase from the seven who mentioned it in the first year. Essentially, five schools progressed from not mentioning value-added data at all to some mention in their plans.

Table 5

Value-Added Data Use in Transformation Plans

	Year #	*N*	%
Extensive use of VA	1	1	7.7
	2	1	7.7
VA used minimally	1	6	46.2
	2	11	84.6
No mention of VA	1	6	46.2
	2	1	7. 7
Total		13	100.00

Note. # = Denotes the academic year of the transformation plan (Year 1 = 2002–2003 & 2003–2004, Year 2 = 2004–2005), VA = Value-added data.

Individual School Results

As Table 6 shows, the overall trends mask considerable variability by school. Analysis for the first-year plans shows total quality scores ranging from 13–22, out of a possible 42 points. For the second-year plans, the range was 15–25. Hoover increased the most (17 points), from 19 in the first rating to 36 in 2003–2004. This school also had the highest total rating of the 13 schools for the 2003–2004 plan. Total scores for Truman and Wilson also increase substantially—13 points and 10 points, respectively. Hoover and Truman are particularly interesting examples of growth, since they went from the bottom one third of scores to the top one-third. Whereas some schools saw minor decreases, most showed at least small increases, although attribution of changes needs to be interpreted cautiously. It is not fully clear whether changes were due to actual changes in data use practices or were a function of communication issues. A school, for example, could have increased their use of data in thinking about their plans but not fully written about the data that influence school thinking (and vice versa).

Table 6

Overall Quality Score by School - Transformation Plan Reviews

School	Year 1	Year 2
Carter	14	24
Fillmore	22	19
Harding	16	19
Hoover	13	25
Jefferson	18	22
Kennedy	20	18
Lincoln	18	19
Pierce	19	19
Polk	17	15
Truman	15	25
Tyler	21	17
Van Buren	19	23
Wilson	13	21

Note. Year 1 = 2002–2003 & 2003–2004, Year 2 = 2004–2005.

Substance of School Plans

Although the transformation plans varied in their use and depth of data, the content of the plans across all 13 schools showed commonalities in their identification of learning goals and instructional strategies.

Identification of school-wide goals. Across all schools, school goals appeared to be derived from a combination of perceived student needs, analysis of test results, and district-wide priorities. School transformation plans revealed, not surprisingly, that most schools focused on improving students' reading, math, or writing skills, often directly written to address problems or weaknesses found in results from annual statewide and or ongoing school testing. Some also included additional goals related to arts instruction, parent involvement, community engagement, or science. Each of the plans also presented broader, more universal school-wide goals (e.g., "academic achievement for all"), which mostly mirrored district-wide directives, but in some cases also incorporates the school's unique sense overarching vision or mission.

Identification of instructional strategies to meet goals. All of the plans outlined in varying degrees of specificity anticipated action steps toward achieving their goals. For example, most schools noted that they would involve their teachers in professional

development programs and or planned to implement specific intervention programs aimed at addressing core goal areas, usually reflecting district wide initiatives.

Frequently cited intervention strategies included literacy intervention programs such as: Sound Partners, Success for All, Read Well, and Six Trait writing strategies: the development of Student Intervention Teams (SIT, targeting individual students needing academic assistance) and Student Academic Plans (SAPs, tailored to individual student needs); differentiated instruction; tutoring and homework help; and technology integration. Most schools also planned to engage their teachers in district-wide professional development in literacy, such as NUA[3] and GLAD[4], which provided instructional strategies that were to be integrated throughout the reading and writing curricula. Additionally, for math, TERC[5] strategies also appeared to be an integral component of most of the plans (TP). In addition to the common strategies, a few schools incorporate strategies as expansive as alignment of curriculum across grade levels, consensus building around a unified philosophy of instruction, and grade-level collaboration to assess student work and design curriculum (TP).

Impact of Transformation Plans

Case study data showed that school transformation plans had varying impact on school operations. In two of the four case study sites (Van Buren and Truman), the great majority of teachers reported that their school's transformation plans guided their classroom planning and had a significant impact on day-to-day practice. Survey responses from the other schools evidences less significant impact, with only 50% of teachers at Polk and only 30% at Wilson reporting impact on curriculum decision making or day-to-day practice.

Schools Use of Data to Inform Instruction and Monitor Progress

Schools mentioned a variety of methods for monitoring progress towards teacher and school-level goals. Almost all of the schools also mentioned the use of curriculum-based assessments (CBAs), and more than half acknowledged the use of portfolios[6]. Teacher-created or classroom-based assessments were reported in over half of the schools' transformation plans, and school-created ones were mentioned in a few as well. Reading,

[3] National Urban Alliance (NUA), a program focused on cognitive development, reasoning, thinking, and higher-order comprehension skills

[4] Guided Language Acquisition Design (GLAD)

[5] Technical Education Research Centers (TERC), an inquiry-based math/science skills program

[6] Explanations of what portfolios encompass were not provided in the transformation plans. However, a primary grade teacher at one of our case study sites (Wilson) stated during an interview that the school has recently gone through the process of developing standardized portfolios across grade levels. With input from staff, a small committee of teachers came up with a proposal for implementing an organized way of collecting students' work (e.g., writing samples) throughout their time at the school. This system is still in the process of being fully implemented by all teachers at this site.

math, and kindergarten inventories as well as teacher anecdotal and observational notes were mentioned in one quarter of the schools' transformation plans. Other types of assessments reported by only a few schools included: Dynamic Indicators of Basic Early Literacy Skills (DIBELS)[7], Multi-Level Skills Inventory-Revised (MASI-R)[8], in addition to the Assess2Learn (an on-line assessment system that scores and analyzes results to help inform instruction).

General Strategies for Monitoring Student Progress

Although virtually all schools reported using assessment data to inform instructional planning and or to adjust instruction, very few specifics were provided in the transformation plans, and based on case study results, relatively few were using assessment on an ongoing basis school-wide. Of the few schools that mentioned specifics in their transformation plans as to how assessment data was used in monitoring student or school progress, schools seemed to focus on some form of pre- and post-testing, using district- or school-based assessments. As an example, Jefferson described that students are assessed on a consistent basis by using school-based pre- and post-tests, to compliment the district and state assessment results. It was unclear, however, whether the same tests were being used at both time points and thus whether improvement could be legitimately inferred. Hoover reported that writing assessments would be conducted at the beginning of each year to establish baseline data to track students' yearly progress. Additionally, Fillmore monitors student academic progress using 8-week reading assessments, as well as quarterly math and DRA assessments, and standards-based Learning System Assessments (TP).

During site visits, principals sometimes were able to provide more detail on progress monitoring, but although teachers generally talked the talk of assessment use, their depth of implementation at the classroom level was moot. For example, the principal at Polk explained that the school had instituted the Assess2Learn program, an online assessment program that "aggregates all the data on your class [and] breaks it down so you really get a feel for what areas classrooms are doing well in." In addition, the principal at Truman recounted that there were many opportunities for students to move to a higher or to a lower reading group throughout the year due to elaborate classroom-based assessments that track students' progress.

[7] A standardized measure of early literacy development used in Kindergarten through 3rd grade.
[8] An assessment for the Read Well curriculum.

Looking Beyond Test Performance

Relatively few schools reported the use of data beyond test scores. From the analysis of transformation plans, a few mentioned reviews of reports or agendas from collaborative teacher meetings (Harding, Tyler, and Wilson), attendance rates in professional development activities (Kennedy, Lincoln, and Tyler), staff survey findings (Harding and Kennedy), and principal observations of teachers (Harding, Jefferson, and Truman). School-level goals, in some cases, were tracked through discussions of progress at Building Leadership Team meetings (Van Buren and Lincoln), School Effectiveness Profile information (Lincoln), and parent surveys or attendance records at school events (Wilson, Kennedy, Polk, Pierce, and Truman). The development of curriculum maps or evidence of curriculum alignment were seen as indicators of teacher professional development or collaboration at Carter, Pierce, and Polk, though these schools did not necessarily describe how alignment was specifically measured. Moreover, it was not fully clear whether and how any of this evidence informed subsequent action.

Other Effects of Testing on Instruction

In addition to examining the extent to which schools and teachers monitored student progress and ostensibly used such evidence for planning curriculum and instruction, we also looked at the effect of direct test preparation (i.e., preparation to do well on state and district tests). Although both ITBS and WASL were routinely administered, WASL got the greater attention. Teacher surveys in the four case study sites addressed how teachers prepared their students for state standards in general and WASL in particular. As Table 7 shows, across the four sites, results indicated that teachers spent more time on addressing the content of state standards, in concept review and writing exercises than on multiple-choice or reading comprehension test preparation exercises. Almost two thirds of the teachers surveyed indicated that they spent *at least one week, if not two or more*, reviewing techniques for organizing ideas into written responses (65%), or reviewing important concepts found in the Washington (WA) content standards (64%). One third of the teachers reported that they spent *two days or less* either discussing and reviewing WASL-type, multiple-choice questions, or reading and answering questions about passages similar to those on WASL. These results are particularly interesting because WASL is only given in 4th grade, but the survey was given to all teachers. Thus, the percentages of teachers spending a great deal of time preparing students for the WASL is much higher than might otherwise be expected.

Table 7

Time Spent on WASL Preparation or Instruction

On average, how much class time did you spend this year preparing for the WASL doing the following activities?	2 days or less	3 to 5 days	1 to 2 weeks	2+ weeks
Reviewing important concepts found in the WA content standards	22%	14%	27%	37%
Discussing and reviewing WASL-type multiple choice questions	33%	17%	21%	29%
Reading and answering questions about passages similar to those on WASL	33%	12%	30%	25%
Reviewing techniques for organizing ideas in written responses	23%	12%	21%	44%

Note. WA = Washington, WASL = Washington Assessment of Student Learning.

Teacher Collaboration and Empowerment Around Assessment Use and Instructional Decision Making

Although systematic data on teacher collaboration and empowerment was not available beyond the four case study sites, the district-wide survey did include an item on the frequency with which teachers met with other teachers to analyze student work. Across the 13 schools, teachers reported that they met approximately once or twice a month. Case study findings confirmed substantial collaboration.

Sharing assessment practices. With curricular and instructional practices, teachers (at the four case-study sites at least) seemed to readily share assessment practices among their colleagues. The CRESST survey inquired as to whether or not teachers had opportunities to share assessments or assessment results, or to align assessment and instructional practices. The majority of teachers surveyed (70%) reported that they shared common classroom-based assessments across grade level. More than two thirds of teachers surveyed (67%) reported that they meet monthly or a few times a year with other faculty and staff to examine the alignment of assessment and curriculum. The frequency with which each of these types of discussions occurred did not vary greatly at each individual site (CS Survey).

Teacher influence over assessment decisions. Teachers across sites often indicated a lack of influence over testing, since the most high-stakes and prominent testing involves state-mandated assessments. However, teachers and principals from at least two sites indicated that teachers do have some influence over other assessment practices. For example, at Wilson, a 2nd-grade teacher described that school-wide assessments, such as reading

assessments, were developed by teachers at the school, "in order to figure out how to serve the needs of the kids."

Support for the use of assessments. Teachers did not report a great deal of support or professional development around the use of assessment results, or even assessment administration. Of the few who mentioned some form of support, a teacher at Wilson remarked that some staff had attended classes on the Direct Reading Assessments that provided instructions for administering the assessment. Additionally, a few teachers at Wilson explained that they had participated in the district's Value-Added project and worked with a district representative to follow a few low-performing students throughout the school year, paying special attention to their ongoing performance on key local assessments and providing additional support when needed. The principal and teachers at Truman reported that professional development and training on assessment data use was provided to teachers. More specifically, teachers were trained in how to interpret and understand the data, and one of the specialists provided some hints on how to improve students' scores based on an analysis of the results.

Usefulness and Accessibility of Data

As schools diversify the types of data they avail themselves of, school staff presumably become more informed consumers of data. The survey and interview data from the four case study sites provides a glimpse into this notion, through teacher and principal perspectives on the usefulness and availability of data. The interview and survey data from these four sites indicates that there is a tension between the availability and usefulness of data. Some teachers expressed a need for more assessments that track individual progress on a more ongoing basis for diagnostic and instructional guidance purposes; on the other hand, teachers are overwhelmed with the amount of data that they must currently sort through. Three major themes emerge from the case study data in regards to the usefulness and accessibility of data:

1. Data is highly accessible, but oftentimes teachers feel overwhelmed by the amount of information they need to sort through;
2. Principals often serve as a data conduit, filtering data for relevancy and helping teachers interpret data, though principals also feel the need for more help in this regard; and
3. The data that schools do have access to is not always the right kind of data and schools have particular criticisms of the data; most prominent in their critique: some school staff want more diagnostic, longitudinal data to help them see individual student growth.

Accessible and Excessive Data

By and large, teachers across the four schools felt that district data are accessible and readily available to them, but often in excess. Almost all teachers surveyed in the four sites (96%) reported that they have ready access to a range of information about school performance (e.g., grade-level assessment results). In interviews, however, teachers consistently expressed feeling inundated with the sheer volume of data they are expected to digest. Some teachers commented that the amount of paperwork and data that they have to deal with is overwhelming and keeps them from doing their job. As one 4th-grade teacher at Polk explained, "I've got reams and reams of data. There is no way I can assimilate all that data. I need some concise data for diagnostic purposes. I'll use it to drive my goals, my teaching objectives, the way I do my planning. That's what I want to see the data used for."

- Nearly all teachers across the four case study sites strongly agreed or agreed that teachers at the school have easy access to individual-level student data (98%), and they also agree that they have adequate information to determine whether a student has mastered a particular topic or subject area (96%) (CS Survey).

Both of these survey items directly relate to the ways in which teachers can use data at the individual student level to provide them with information on students' current level and then determine how best to support their learning processes to help them succeed.

Principal as Data Conduit

Perhaps in response to teachers feeling overwhelmed by the amount of data involved, principals at three of the four schools report serving as data conduits, doing their best to interpret data for their teachers, or to provide avenues for healthy data discussions (with the exception of Polk, where the principal indicated that he was struggling himself with data analysis and use). Despite their own struggles with the usability of the data they receive, principals at the four case study sites are making great efforts to encourage, support, and increase teachers' data use.

At Van Buren, the principal "translates" and summarizes data and puts the data "in front of them all the time." She explained how she streamlines the substantial amount of assessment data she receives from the district to make it more manageable for her teachers. She also helps to translate the value-added data for her teachers, whom she says tend to discount it, as they do not see it as reflective of their students' actual progress. The principal at Wilson explained that she finds district data to be a very useful tool for informing instruction when she "can get through it all." In fact, she described how she has enlisted the school's NUA consultant to help her sift through the data and interpret it, further noting that

she ultimately determines which data is most useful in commenting, "When I make [the data] relevant, then it is." She analyzes district data and disperses the most relevant components to her teachers, and she also commented that she feels that each year they are able to look at the data in more depth. Teachers at the school are also required to set personal academic goals, and student data are used to assess the degree to which they have achieved these goals.

Interview data from Truman suggests that the principal encourages data use amongst her teachers as well. She reported that she and the reading coach attend a monthly meeting with other schools to share ideas for implementing strategies to assess students and to score the assessments so that teachers can look at the results immediately. As the teachers learn to assess, score, and analyze the data, the principal sets expectations surrounding these data with her teachers, and asks teachers to bring in data for the pre-conference. As the academic year progresses, the principal asks the teachers to discuss their students' growth since the first assessment at the beginning of the year. According to the principal, one teacher shared the following with the principal: "You know, when we first looked at this 2 months ago, I had seven kids up here in the risk area. I moved them to the reader area because they've done so well." The principal commented, "So it's finding more simplistic ways for folks to use data that's really useful." Staff meetings are also used as an opportunity to discuss data, where visual displays of progress are often used on either district or school-based assessments.

Unlike the other three case study schools, the depth and consistency of data use at Polk was much less clear, and the principal was very candid in stating his own need for help in interpreting district data. He stated that he (and in turn, the school as a whole) could greatly benefit from training and on-site support in data analysis. A number of teachers repeated the principal's suggestion for more support in interpreting the district data, because they did not feel that they were using the data to their students' best advantage. Although this need for support was most clear at Polk, teachers at all of the case study sites mentioned that they could benefit from increased support in interpreting and using district data.

Some Data Less Relevant to Practice

Teachers are more divided, however, about the relevance and usefulness of standardized test data to their practice. Since the most prominent data cited by schools as evidence of progress are the state and district assessment results, case study schools were asked how helpful these data are to their practice. Across the four schools, half of the teachers surveyed agreed or strongly agreed that district data sources are usually *not* helpful to them in planning instruction. This division of opinion regarding the relevance and

usefulness of district data was also reflected in the interview data for the four case study schools. Teacher feedback across school sites indicated that teachers have some doubts about the usefulness of district data. A few teachers questioned whether district-level data is truly reflective of student progress, often remarking that it does not present the "whole picture" of individual students, and (by not providing pre–post information) does not allow teachers to effectively track individual student progress. The sources of frustration with the relevance of state testing data indicate a growing sophistication in knowledge of data, and include:

- *Lack of Continuity in Tests, or Longitudinal Data*: Teachers and principals across the four sites expressed frustrations that their largest data source (e.g., WASL and ITBS scores) compares different groups of students to each other (consecutive 3rd-grade classes), as well as the same group of students on different tests (value-added analyses tracking growth from ITBS to WASL). When trying to effectively use the data to inform instruction, teachers expressed a need to have individual students tracked from year to year on the same assessment, so that progress can be monitored. Primary-grade teachers at Van Buren, for example, suggested that if there were more consistency with the type of assessments used, this would allow time to see the effect of their efforts on student growth. A few teachers felt the change in test (ITBS and WASL) from grade to grade made it more difficult for them to standardize interpretations of students' results. The fact that the value-added data looks at data from two different tests makes teachers suspect as well. As one teacher described, "There's a quality about value-added data, which is a little bit like believing in magic. Even though people stand up there and seem to believe it and understand it, there is a part that just feels like you are comparing apples and oranges. I think this is true of a lot of data."
- *Lack of Diagnostic Information*: Still other teachers noted that the much of the state testing data does not provide them with useful information beyond broad brush measures of what they already know about their students' skill levels. A number of teachers remarked that teacher-created, classroom-based assessments are much more effective in measuring student skills and progress, that these tests give them a clearer picture of how individual students are doing and how to adjust instruction to best meet their needs. As one lower-grade elementary teacher at Truman explained, "With the Read Well assessments it's pretty black-and-white. You see their progress. As far as the other district assessments, it's been kind of a learning process as far as what it all means." Teachers at Wilson, on the other hand, pointed to their disappointment at potentially losing the DWA at their site next year, because the DWA is one assessment that they identified as providing diagnostic information on students' progress over time.
- *Appropriateness of the Assessment:* Some teachers expressed the more general sentiment that standardized testing, particularly the WASL, does not take into consideration variation in developmental readiness among students, particularly for students with special needs. As one principal noted (at Truman): "First we've got to make sure we've got an instrument that's useful and that's focused and that can be broadly used and interpreted. I don't think that we'll find an assessment that is

going to be totally culturally relevant or any of that. But, whatever tool we decide to use, it needs to be one that we can within our spirits and souls say, 'This really is reflective of children, first through fifth grade or first through twelfth grade, no matter where they live in the city.'" This concern about the appropriateness of the assessment for students with special needs was prevalent across interview data for the four sites, but was most marked at Polk.

- *Provides a Narrow Picture of Student Performance—Not the Whole Picture:* Teachers at all four schools offered a criticism of WASL and ITBS data and standardized tests in general, explaining that they felt the resulting information does not capture the "whole picture" of student capability and progress. Teachers questioned how well the tests reflect students' knowledge in general. Some teachers cited concerns regarding cultural relevance of standardized test scores. Although teachers across the four schools commented on this theme, it was clearly most pronounced at Polk, which serves a very low SES population and is, as the principal explained, struggling in terms of standardized test scores.
- *Focusing on Status Rather than Growth:* Many teachers described their frustration with the district's measure of student success as meeting a static, absolute performance standard on standardized tests, rather than taking into account individual student growth over time. As an example of the sentiment felt by many teachers, one 4th-grade teacher explained the importance of taking into account where a student starts out and grows when considering progress:

> The problem that I have with the standardized testing is it measures a destination. It doesn't measure a journey. A kid can have traveled this far (my gesture is a very short distance), and have made standards. The district's saying, 'Whoo-hoo, you guys are doing a great job.' Yes, the destination is important, but there's got to be some value in the journey these kids have made. Last year in WASLs, we had ten to twelve that made eights rather than nines. They were so close. They don't get to hear, 'Good job. You're so close.' All they get is, 'Not at standard.'

Another 4th-grade teacher gave the example of a specific student's progress to illustrate the same point:

> She started out at a low kindergarten reading level. Now she's reading at a high second grade, actually at a low third grade level. That's huge. Three years of progress, but she's not at standard so it won't show up. Not only does it not show up for her, but it doesn't show up for us. That gets kind of frustrating.

- *High Stakes Nature of State Testing Makes Results Punitive*: Some teachers further suggested that state testing, because it is primarily used for accountability purposes, is frequently used punitively for both students and schools. For example, the principal and 4th-grade teacher at Polk explained how WASL scores have been used negatively against the school, through public stigma, as opposed to driving improvements in instruction:

> This year was my first experience other than as a parent with [WASL]. The higher-level thinking, I'm all for … The problem is when something that could give you useful information is used … not only against you, but to be the be-all and end- all to that child and that teacher, that principal and that school. They keep saying that when the WASL was first developed it was designed to help drive instruction, and show how and what [we need] in this area … As soon as it becomes published in the paper who fails, who doesn't, [the message becomes] 'Don't go to that school they've got terrible test scores.' The emphasis is totally wrong … So it's punitive.

- *Usefulness of Value-Added Data:* Principals and teachers at several of the four case study sites expressed concern over the relevance or usefulness of the value-added data, in particular. Some sites critiqued the fact that value-added data is not provided to them at the individual level or disaggregated by ethnic group (some intermediate teachers at Van Buren particularly emphasized the need for ethnic data since they serve an extremely diverse population). Teachers expressed the same frustrations as with other statewide testing results—they have a hard time interpreting and analyzing the results, so do not always find the time, or have to rely on principals or others to interpret the data for them. Also, because of the change in tests mentioned before, some teachers are highly skeptical of the data and what it represents.
- Wilson stood out among the four case study sites as a school that seemed to have a slightly stronger sense of how value-added data works, and interviews suggested that value-added data are a more integral part of this school's process. Teachers were given the opportunity to work with a district representative (value-added coach) to look more closely at the value-added data and follow a few individual students to monitor their progress over time. Two teachers who volunteered for this experience explained that the process of monitoring students through observational notes, work samples, and pre- and post-test data allowed to better define students' needs, and positively impacted her instructional planning. Other teachers also described a collaborative process by which a group of teachers examined value-added data to understand the discrepancy between student classroom performance and their scores on statewide testing programs. This collaborative analysis process led to some changes in school wide instructional decisions. Despite the positive feedback from some teachers at Wilson about the value-added data, some staff also questioned its bearing on their practice. Members of the BLT cited a specific example when trying to disaggregate data to identify particular students for intervention where the individual students who needed further intervention could not be individually identified (requests to the district for clarification and identification of students were fruitless).

These critiques of the assessments used and the data reported resulted in many explicit suggestions made by schools about how to improve data use, as well as some implicit recommendations suggested by the types of critiques that emerged. These can be found in the Conclusions section at the end of this report.

The Overall Story of Data Use from Case Study Sites

From our case study data, schools report that they have access to a great deal of data and are using it for reform and instructional planning purposes. Nearly all teachers (92%) reported that their respective schools carefully examine district-provided test results on an annual basis, and a majority of teachers (60%) reported that grade-level test results guide their instructional planning (CS Survey). This information indicates that schools are consistent about the fact that they review test data annually, but are less sanguine about the degree to which they use it for instructional planning purposes. In this study, we attempted to separate out the somewhat perfunctory annual review of data for informational purposes from the use of data to inform instructional planning decisions on an ongoing or more periodic basis. To the extent to which it was available, information was gathered from schools' transformation plans—and buttressed by case study interviews and surveys—about the degree to which data is used for school-wide planning or improvement efforts or instructional change at a grade level or individual classroom level and the process for reviewing individual student progress.

Data Use Informing School-wide Transformation Planning & Improvement

Although most schools did not explicitly state the process by which transformation plan goals were identified, some examples were included in several of the school's plans. To illustrate: (a) two schools (Carter and Kennedy) conducted needs assessments to identify areas for improvement; (b) at another school (Fillmore), staff focus groups used information from longitudinal assessment data and current research to identify four areas for improvement; and (c) three schools described that collaborative processes, led by the schools' Building Leadership Team were used to form the schools' goals (Fillmore, Lincoln, and Pierce). In other words, half of the schools explicitly used a collaborative process, with at least three even more explicitly using data to drive the discussions. Case study interviews and surveys appear to confirm the collaborative aspects of progress monitoring, and even strengthen the link between data and the transformation planning process. For example, 86% of teachers at all four case study sites strongly agreed or agreed that teachers at the school play an active role in monitoring progress toward the schools' goals (CS Survey). And, when areas of weakness (e.g., math measurement) were identified by data sources such as WASL scores or item analysis on ITBS, three of the four schools (Van Buren, Wilson, and Truman) consistently tended to focus attention on that particular topic in their school-wide goals (TP and CS Interviews).

Three of the four case study sites identified methods by which they monitor progress towards school-wide improvement goals (Van Buren, Wilson, and Truman). Although each of these schools indicated that positive changes in student learning and work products (classroom-based and other more formalized assessment results) was indicative of progress, all three described some unique, and specific ways that they look for signs of school-wide progress. For example, at Wilson and Truman, they collect baseline information from a particular assessment and then gather comparative information throughout the school year to see student growth, and use teachers' observations of student engagement in learning as well (CS Interviews). When asked about the frequency of using student assessments to evaluate teaching effectiveness, responses were spread throughout the scale with 62% of teachers acknowledging that this occurs once a month or more (CS Interviews and CS Survey). The split in agreement about the frequency may suggest that certain schools and grade levels are using assessments to evaluate teaching more frequently than others. This amount of agreement as it is, however, seems quite high, since a once a month review of data to evaluate teaching effectiveness is fairly frequent.

Data Use Informing Instructional Change

Use of data for instructional change at the classroom level or at a grade level was not as frequent or as consistent across sites, but still reportedly quite high.

- Altogether, 76% of survey respondents at all four sites strongly agree or agree that they use results from classroom-based assessments when deciding how to make changes to instruction. Furthermore, in planning lessons, only about half of teachers from all four sites (49%) reported that they rely on feedback from classroom-based student assessments *a great deal* (the scale ranged from *not at all* to *a great deal*, CS Survey).
- Forty-three percent of teachers across all sites responded that they rely on external examinations or standardized tests "some" of the time when planning lessons. This result flows with interview data as even though most teachers described that standardized testing data is considered and impacts what they do, they seem resistant to rely too heavily on this data knowing that there are problems with the data and that it is difficult to use in a diagnostic manner (CS Interviews and CS Survey).
- Across the four schools, 67% of teachers acknowledged that they meet with other teachers to discuss using student assessments to tailor instruction to students' skill levels once a month or more (CS Survey).

Data Use to Monitor Individual Student Needs and Progress

Another critical piece that incorporates data use practices and monitoring is the method by which schools identify low-performing students in need of additional support. No

systematic data across all 13 sites was available from transformation plans on this level of specificity in data use practices. All four schools, in varying ways, report that when possible they interpret data at the individual student level and use this information to identify strategies to help these students become more successful learners. As mentioned earlier, almost all teachers responding to the survey in our four case study schools felt they had easy access to individual-level student data (98%), and had adequate information to determine whether a student has mastered a particular topic or subject area (96%, CS Survey). This contrasts with the interview data on usefulness and relevance for monitoring student progress, where much of the data seemed to suggest that the assessments that teachers have access to now do not provide this type of individually based diagnostic and formative information. Perhaps teachers and principals assumed that the interviewers were asking only about statewide testing, when they responded so strongly that the testing information they have access to was not diagnostic or individual enough.

For Polk, identifying individual student progress is the major focus of their data use. Wilson uses school-wide assemblies to report on individual students' progress toward proficiency on certain math or reading (fluency) skills. The other two schools use data for broader purposes, such as tracking grade-level or school-wide goals. All four schools cited the use of Student Intervention Teams (SIT) to identify students with particular learning challenges, and then created a plan that will be used to help a student improve (CS Interviews).

Process or Mechanism and Support for Using Data

The degree to which schools use a collaborative process to review data varies. As mentioned earlier, the principal is often the conduit of information and data, and serves to interpret and filter much of the data on behalf of teachers. As our Transformation Plan Rating process indicated across the 13 schools, most used a fairly standard model in terms of who was included in the process of reviewing data and monitoring progress. Most schools had at least the minimal amount of inclusiveness of school community stakeholders, with some mention of the BLT, PTA, and other common groups; but very few included systematic involvement of teacher groups or grade-level teams in the analysis of progress or decision making.

Collaborative Process Involving Teachers. Case study data confirmed that schools used a collaborative and inclusive process to discuss data. The principal, by and large, drove this process, with often a great deal of teacher involvement (less teacher involvement at Van Buren than at Wilson and Truman sites). Polk, as mentioned before, was in need of support for the principal and the teachers in interpreting and analyzing data.

- As described above, over two thirds of teachers across the four case study sites (67%) acknowledged that they meet with other teachers to discuss using student assessments to tailor instruction to students' skill levels once a month or more, and there was little variation between sites. This data supports interview data that teachers are collaborating on a regular basis and discussing student data and how to modify instruction based on findings. Furthermore, 88% of survey respondents noted that once a month or more they meet with other teachers to discuss student progress and learning (CS Interviews and CS Survey).
- As one explicit and extensive example, Truman provides opportunities for teachers to consider data during staff meetings every first and third week of each month. During these meetings, data, which often include literacy and mathematics information from district assessments, are shared among teachers. One of the specialists reported that both classroom and grade-level data is examined during this time. She stated, "The test that I'm doing now, when we get the results back, I bring it back to the teachers and we look at it. The way it's set up, you can visually see the areas that we need to work on and the areas we're being successful in."

Data review and monitoring occurring frequently. Schools, on the whole, tended to improve the frequency with which they analyzed data or monitored progress; on average, schools are definitely monitoring progress more than annually, though how much more than annually we do not know. Though review is occurring more frequently, schools do not yet seem to have the data review or monitoring process fully integrated into their decision making process (TP Ratings).

- A data review process, primarily of student assessment results, was described explicitly in the transformation plans at five schools (Carter, Hoover, Fillmore, Lincoln, and Pierce). In these schools, various approaches, including review of student data by all staff in the fall and monthly grade-level meetings to analyze student academic achievement patterns, are used to monitor progress (TP). For example, fall early release days are used (at Carter) for the entire staff to review student assessment data and develop grade-level, and classroom-specific goals to support student learning. At Kennedy, data from 1999–2001 were analyzed by staff to look at student academic achievement patterns and the relative effect of instruction and curricular programs. Moreover, aside from the regular monitoring of student assessments, nearly all schools mentioned that Student Intervention Teams (SIT) were used to review individual students' progress for students with particular academic, social, or emotional concerns.
- In two of the four case study schools (Wilson and Polk) 100% of teachers said dialogue about student progress and learning was happening once a month or more, and at Wilson, 84% of teachers acknowledged discussing student progress and learning once a month or more. Van Buren had the lowest agreement about the frequency of tracking individual progress, but still had a large majority (73%) of teachers reporting that this occurs once a month or more. The results seem to indicate that teachers across all four schools are largely in agreement that review of

individual student progress is a frequent occurrence, at least once a month if not more frequently.

- The district staff climate survey asks a handful of questions to all schools that are related to data use. The opinion scale asked teachers the extent to which they agree or disagree with a set of questions such as: availability of useful information to make informed decisions, the school's stance towards their work being one of inquiry and reflection, assessment of student performance leading to curricular change, and teacher engagement in systematic analysis of student performance data. Across the district, the average endorsement of these items was 3.85, indicating that teachers across the district, on average, consistently agree that their school's practices reflect the data use items (5-point scale, 1 (*strongly disagree*), 5 (*strongly agree*).[9] Across the 13 schools in our sample, the average response to these items was 3.81, indicating that these schools also consistently agree that these practices or characteristics are reflected in their schools. In addition, the district survey asked about the frequency of one particular data use practice (e.g., analyzing student work with other teachers). The frequency scale was also a 5-point scale: 1 (*never*), 2 (*few times a year*), 3 (*once or twice a month*), 4 (*once or twice a week*), 5 (*almost daily*). The district average frequency on this item was 2.61, whereas the average across the 13 schools was 2.72, again indicating similar patterns between the 13 schools in this study and the district on a whole. Overall, it appears that teachers believe they are analyzing student work somewhere between a few times a year and once or twice a month. This variation could be a result of different patterns between grade levels (primary vs. upper grades).
- In general, there was little variation in each school's mean scores, though some distinctions surfaced. With regard to the data use or inquiry opinion scales, two schools, Kennedy and Tyler scored significantly below the district average on the opinion scale (agree or disagree) and only one school (Lincoln) scored significantly below the district average for the frequency of analyzing student work. Furthermore, in examining individual schools' average scores over time, from 2001–2004, two schools' (Jefferson and Polk) scores on the opinion scale increased by one point or more. These schools are reaching stronger agreement over time about their schools' data use practices. On the frequency of behavior scale over this period of time, one school's score decreased by one point or more since 2001 (Wilson), indicating that teachers, on average, feel that they are analyzing student work less frequently in recent years. This result may at first glance appear to conflict with case study data, in that Wilson was one of the four sites that appeared to be using data most frequently and had a stronger and widespread understanding of how to interpret data. However, because the question is so narrowly limited to the analysis of student work (only one form of data analysis), we do not believe that the information is truly contradictory.

[9] We are aware that it is not always prudent to average scales that are not interval scales; however, for these purposes, we have done so, in order to capture overall trends across the district and across the schools in our sample. In addition, we are using these averages to get a sense of any strong outliers, so we are only using the averages to capture broad-brush patterns, not to distinguish subtle or slight distinctions between schools.

Achievement Results

The 13 schools in this study were initially selected based on their differential growth across two time points, as students moved from 3rd grade to 5th grade. We looked at stability across two cohorts. Specifically, seven schools were identified that started in 2001 with below average 3rd-grade ITBS scores, but whose students by 5th grade had shown greater-than-average growth relative to the district for the school as a whole and or for their initially low performing students and that had shown similar patterns of growth from 1999 to 2001. These schools were termed Break the Mold schools. Six demographically similar schools were selected as comparison sites. We explored these schools to learn more about their data use practices, and to examine whether these practices seemed to lead to different patterns in growth, according to our measures. As we reviewed transformation plans, conducted site visits and analyzed our cases, blind to which schools were identified as Break the Mold, we also accessed additional student data so that we could assess growth over additional cohorts and confirm schools' "effective" status over the period of data collection. These analyses ended up raising more questions than answers, and illuminated endemic difficulties in identifying a stable set of effective schools based on student performance, despite the application of advanced statistical methodologies.

CRESST Value-Added Model of Growth

As mentioned in the methods section, we conducted latent variable, multilevel analyses to estimate the average gains experienced by performance subgroups in each school in reading and math on the ITBS, using student-level, longitudinal data and estimated average gains (or value-added) across two time points (3rd grade to 5th grade) on the ITBS for each successive cohort. (See Methods section for more details.)

Our subgroups are defined by their initial performance status rather than by their ethnic, language, disability, or income levels. Instead, we analyze the distribution of schools' added value across performance subgroups. For a broader perspective on growth, we analyzed the average growth trajectories for three groups:

- Average: students who started at the mean initial status;
- Low: students who started at 15 points below the school mean initial status; and
- High: students who started at 15 points above the school mean initial status.

Because expected mean gains may provide an incomplete picture of whether schools may be leaving some children behind, we also examined the expected gains by different levels of student initial achievement status. Furthermore, over the course of the study, we

expand this longitudinal analysis further by tracking five contiguous cohorts longitudinally. These cohorts begin with the 1998 cohort, where students were in the 3rd grade in 1998 and 5th grade in 2000. The last cohort included was the 2002 cohort, where students were in 3rd grade in 2002 and 5th grade in 2004. More specifically, the cohorts—and the gains experienced by these cohorts—as referred to throughout this section are as follows:

- Cohort 1 (1998): growth of students from 3rd grade in 1998 to 5th grade in 2000;
- Cohort 2 (1999): growth of students from 3rd grade in 1999 to 5th grade in 2001;
- Cohort 3 (2000): growth of students from 3rd grade in 2000 to 5th grade in 2002;
- Cohort 4 (2001): growth of students from 3rd grade in 2001 to 5th grade in 2003; and
- Cohort 5 (2002): growth of students from 3rd grade in 2002 to 5th grade in 2004.

Once the low, average, and high gains were estimated for each school, and for each of the five successive cohorts, our analytic approach focused on comparing overall reading and math performance for different sets of cohorts. We divided cohorts by whether they would have been influenced by the transformation planning process (or the district's focus on data use), which began in earnest during the latter part of the 2001–2002 school year. Gains in achievement levels experienced by Cohorts 1–3 would presumably not be influenced by the transformation planning process since testing—and, more importantly, the instruction that would impact testing—for those cohorts would have occurred in time periods preceding the bulk of the transformation planning efforts. These cohorts represent our *pre-reform* group. Cohort 5, whose 3rd graders started in the 2001–2002 school year, would, to the contrary, have experienced the bulk of the transformation planning process. Cohort 5 represents our *post-reform* group. Because Cohort 4 straddles this time period (3rd graders start in 2000-2001 school year and are in 5th grade by 2002-2003 school year), we did not use them in the comparisons shown in Table 8.

The *difference column* in Table 8 shows the difference between the pre-reform cohort gains (averaged across the three cohorts) and the post-reform cohort, which unfortunately identifies a different group of effective schools from our initial analyses. As the reference point, district differences in gains between these two groups are close to zero in reading and slightly negative in math. The 13 schools in our sample, on the other hand, demonstrate slightly positive difference scores, though math difference scores are negative. Important to note, however, is the trend for the below-average subgroup. Below-average students tend to make more positive gains from pre- to post-reform in reading, and the least negative gains from pre- to post-reform in math. This trend suggests that perhaps the transformation planning process may help schools—in some small way—focus more attention on the low performers in their schools.

Table 8

Comparisons of Pre and Post Math and Reading Gains: Low, Average, and High Subgroups

		READING			MATH		
Name	Subgroup	Pre (Coh 1–3)	Post (Coh 5)	Diff	Pre (Coh 1–3)	Post (Coh 5)	Diff
Van Buren	Low	25.24	26.42	1.18	28.09	20.24	-7.85
	Average	27.78	31.44	3.66	32.99	21.68	-11.31
	High	30.33	36.45	6.12	37.88	23.12	-14.76
Carter	Low	27.56	26.66	-0.90	24.49	25.91	1.42
	Average	30.01	27.48	-2.53	27.94	26.07	-1.87
	High	32.46	28.31	-4.15	31.39	26.23	-5.16
Harding	Low	25.36	38.76	13.40	24.46	27.26	2.80
	Average	33.49	36.34	2.85	32.43	23.83	-8.60
	High	41.62	33.93	-7.69	40.41	20.41	-20.00
Hoover	Low	35.97	38.84	2.87	37.68	42.76	5.08
	Average	37.62	37.82	0.20	41.75	41.44	-0.31
	High	39.28	36.80	-2.48	45.81	40.12	-5.69
Fillmore	Low	23.79	36.74	12.95	21.68	21.79	0.11
	Average	28.25	32.91	4.66	26.82	21.24	-5.58
	High	32.71	29.09	-3.62	31.96	20.70	-11.26
Polk	Low	22.60	37.79	15.19	20.12	22.60	2.48
	Average	29.26	35.72	6.46	21.46	24.54	3.08
	High	35.92	33.66	-2.26	22.80	26.48	3.68
Kennedy	Low	30.39	28.48	-1.91	33.57	23.70	-9.87
	Average	30.99	30.39	-0.60	36.54	27.64	-8.90
	High	31.60	32.30	0.70	39.49	31.58	-7.91
Jefferson	Low	24.26	20.19	-4.07	18.52	22.16	3.64
	Average	22.86	27.57	4.71	20.77	25.39	4.62
	High	21.46	34.95	13.49	23.01	28.62	5.61
Pierce	Low	25.35	17.93	-7.42	32.13	19.73	-12.40
	Average	27.31	22.32	-4.99	31.41	18.95	-12.46
	High	29.26	26.70	-2.56	30.69	18.17	-12.52
Truman	Low	23.65	31.05	7.40	13.74	25.72	11.98
	Average	22.95	32.79	9.84	14.79	26.79	12.00
	High	22.24	34.53	12.29	15.84	27.86	12.02

(*table continues*)

Table 8 (continued)

Name	Subgroup	READING Pre (Coh 1–3)	READING Post (Coh 5)	READING Diff	MATH Pre (Coh 1–3)	MATH Post (Coh 5)	MATH Diff
Lincoln	Low	18.95	12.54	-6.41	19.79	9.66	-10.13
	Average	22.80	20.39	-2.41	25.74	14.30	-11.44
	High	26.65	28.25	1.60	31.69	18.94	-12.75
Tyler	Low	26.63	23.37	-3.26	27.01	27.30	0.29
	Average	30.73	30.05	-0.68	30.24	31.04	0.80
	High	34.83	36.72	1.89	33.47	34.78	1.31
Wilson	Low	27.86	29.42	1.56	26.79	23.05	-3.74
	Average	32.23	33.91	1.68	29.32	26.10	-3.22
	High	36.61	38.41	1.80	31.85	29.15	-2.70
13 schools	Low	25.97	28.32	2.35	25.24	23.99	-1.25
	Average	28.95	30.70	1.76	28.63	25.31	-3.32
	High	31.92	33.08	1.16	32.02	26.63	-5.39
District	Average	29.55	29.00	-0.55	30.87	27.71	-3.16

Note. Coh = cohort

The patterns we found in the gains and difference scores among our 13 schools are further synthesized in Table 9 and explained below. We describe these different patterns in terms of high or average growth, low growth, or very low growth. It is important to note that in the analytic table above, schools are being compared against average change in performance for the sample from pre- to post-reform. This means that schools' classifications of performance below are, in very real ways, sample dependent (e.g., if we were to do this analysis again with a different set of schools, the results might be different). In other words, it is important to remember that this is a purposive sample, with schools chosen based on certain background and performance characteristics, e.g., above the district average percentage of low-SES students and below-average 3rd-grade ITBS scores and represented a group of schools that were already struggling with student achievement. We analyzed the pattern of gains pre-transformation planning and post, to see if there were differences among the schools in terms of growth from their initial position of achievement difficulty. Moreover, the data in Table 9 show the difficulty of finding stable patterns of performance across both reading and math and across initial status in achievement levels.

Table 9

Growth Patterns Compared to Sample Average

School	Reading	Math	Overall
Truman	>> avg hi; > avg all	>> avg all	High growth
Polk	>> avg low	> avg all	High growth
Wilson	avg	avg	Average growth
Hoover	< avg hi	> avg low	Average growth
Jefferson	>> avg hi; < avg low	> avg all	Low growth (ach gap: rdg; good growth in math)
Tyler	< avg low	> avg all	Low growth (low growth for students starting low in rdg; good growth in math)
Van Buren	> avg hi	< avg all	Low growth (low growth in math)
Carter	< avg all	Avg	Low growth (low growth in rdg)
Harding	>> avg low; < avg hi	<<avg hi; > avg low	Low growth (ach gap: rdg and math)
Fillmore	>>avg low; < avg hi	<< avg hi	Low growth (ach gap: rdg and math)
Kennedy	< avg low	<< avg low	Very low growth
Lincoln	< avg low	<< avg all	Very low growth
Pierce	< avg all	<< avg all	Very low growth

Note. avg = average, ach = achievement, rdg = reading gains.

High growth or average growth. Four schools fell into the high growth or average growth categories (two in each)—Polk and Truman (high growth) and Hoover and Wilson (average growth). These schools, for the most part, showed consistent growth from the pre- to post-reform periods across all three performance subgroups—students starting out in the middle, students starting out above the school average, and students starting out below the school average. Truman had particular success in improving gains for students who started out above the school average in reading; but also had surpassing success in improving gains for all students in math. Polk had particular success in improving gains for students starting out below the school's average in reading. Hoover is rated as showing average growth between pre- and post-reform periods because, though the students who started out high in reading showed negative gains, it was only slight; whereas students who started out low in

math showed largely positive gains (compared with the sample average, which was negative).

Low growth or leaving subgroups behind. Low growth is the largest category in Table 9 (six schools), and indicates that the schools saw less than average change in their math and or reading gains from pre- to post-reform, compared with the sample as a whole. When there was a marked difference between students starting out high and students starting out low, this was also considered a low-growth school, since that disparity—or gap in achievement—has been targeted by the district as something to be eradicated. If a school experienced low growth in one subject but high growth in the other, they were still considered a low-growth school. This decision mimics the rigid nature of Adequate Yearly Progress (AYP), which requires that all subgroups must make AYP in both reading and math, or the school will not make AYP. Schools in this category are experiencing difficulty making gains with particular subgroups (either the students starting out low, or the students starting out high) or in particular subject areas (reading or math), but are **not** necessarily consistently showing little to no gains as with the next category.

The six schools in this category—Jefferson, Tyler, Van Buren, Carter, Harding, and Fillmore—have very different patterns of growth among them. Jefferson and Tyler are arguably the better performers of the bunch, since they have good, consistent growth in math for all subgroups, but showed less-than-average growth for students starting out low (Tyler) or showed a gap between high and low students in reading (Jefferson had far greater than average gains for students starting out *above* average at the school, but less-than-average gains for students starting out *below* average).

Van Buren and Carter, however, had fairly average growth in one subject area, but made less-than-average gains for *all* subgroups in the other (Van Buren left all students behind in *math* and Carter in *reading*). Harding and Fillmore had achievement gaps in both subject areas between students starting out *above* and *below* average. Harding appears to make greater gains with students starting out below average (which is a commendable thing), since this group was making far greater than average gains in reading between the pre- and post-reform periods, whereas students starting out above average showed lower-than-average gains. In math, Harding students starting out below average showed greater-than-average gains, but students starting out above average showed far lower than average gains. Fillmore, on the other hand, appears to show far greater gains for students starting out above average at their school in reading, but far worse for these same students in math.

Very low growth. This group of three schools (Kennedy, Lincoln, and Pierce) had consistent patterns of no growth or negative growth (decline in gains) from pre- to post-reform periods for either for all students, or just for students starting out below the schools' average. Kennedy is having a particular hard time showing any gains for the students starting out below average in both reading and math, while Lincoln is leaving *all* students *far* behind in *math* and only the students starting out below average behind in *reading*, and Pierce is leaving *all* students *far* behind in *math* and behind in *reading*.

Achievement and Data Use

When matching these patterns of growth between pre- and post-transformation planning to the indicators we have of data use from the transformation plans themselves, we must return to the data use ratings discussed in previous sections. As Table 10 indicates, many schools (10) are consistent in terms of their absolute ratings of high, medium, and low as well as the improvements that were observed from the first analysis of transformation plans to the second. Three schools cross over categories between their absolute and improvement ratings: Van Buren, Fillmore, and Pierce. For each, decisions had to be made to place them into only one category. Van Buren had high overall scores and made four points of gains. Because they started and remained as one of the higher scoring schools overall, we placed them in the high category, even though they showed only modest improvement. Pierce and Fillmore both had middle range absolute scores, but showed no growth (Pierce) or negative growth (Fillmore). Because these schools made no improvement or even fell behind on their ratings, we chose to put them both in the low category.

Table 10

Transformation Plan Ratings: Data Use

	High	Medium	Low
Absolute Ratings	*N* = 5 Scored 30 points or more (out of 42): Van Buren, Carter, Hoover, Truman, Wilson	*N* = 5 Scored 25–29 points: Harding, Fillmore, Jefferson, Lincoln, Pierce	*N* = 3 Scored < 25 points: Kennedy, POLK, Tyler
TP Ratings Improvement	*N* = 4 Improved quality ratings by 8–12 points: Carter (#5) Hoover (#7) Truman (#40) Wilson (#17) (Hoover and Truman went from bottom 1/3 to top 1/3)	*N* = 4 Improved ratings by 1–4 points: Van Buren (#2) Harding (#6) Jefferson (#27) Lincoln (#62)	*N* = 5 No improvement or decreased ratings (negative gains): Fillmore (#18) Kennedy (#26) Pierce (#33) POLK (#23) Tyler (#37)
Overall	*N* = 5 Carter, Hoover, Truman, Wilson, Van Buren	*N* = 3 Harding, Jefferson, Lincoln	*N* = 5 Kennedy, POLK, Tyler, Fillmore, Pierce

As Table 11 shows, for some schools there are consistencies between their data use ratings and their pre- and post-transformation gains in student performance. Seven of the 13 schools share the same high/medium/low rating the two classifications. Yet, the data also show important discrepancies between data use and performance. Polk provides an extreme example: while the school shows consistently high growth in student performance, and is one of the top performers in our gains analyses, it was consistently judged a low performer in terms of data use. The site visit data showed a school in flux and one that perceives itself to be a low performer, despite the CRESST analyses. The principal had been there only a few years and seemed overwhelmed with the pressure that the district has put on him to turn around a historically low-performing school in a short amount of time. While data use at the school was limited, the principal did speak of establishing and enforcing new expectations for staff and student attendance, regular schedules, and abolishing multiple-grade classrooms. The principal spoke of a school in chaos when he first arrived, where teachers were frequently absent, substitutes could not be or were not found, so classes were often combined. It appears that creating basic standards and routines may be having an effect and the problems are such that the school does not yet have the time or energy to devote to data use practices.

Table 11

Pre-Post Gains Compared with Data Use Practices

School	Pre–Post Differences	Data Use Practices
Truman	High growth	High
Polk	High growth	Low
Wilson	Average growth	High
Hoover	Average growth	High
Jefferson	Low growth	Medium
Tyler	Low growth	Low
Van Buren	Low growth	High
Carter	Low growth	High
Harding	Low growth	Medium
Fillmore	Low growth	Low
Kennedy	Very low growth	Low
Lincoln	Very low Growth	Medium
Pierce	Very low Growth	Low

Other discrepancies include: Tyler, Van Buren, Carter, Fillmore, and Lincoln. The only other school for which we have site visit data to help explain some of these discrepancies is Van Buren. Van Buren was found to have one of the higher scores overall on their data use practice ratings, but also appeared to have consistent problems in raising math achievement for all subgroups. While site visit data confirmed that the school does, in fact, appear to be very data-driven—with the principal helping teachers access instructionally-relevant data for individual students (see Data Use Vignettes, Appendix C)—site visit data also revealed that the interventions and the reforms at the school seem to focus primarily on reading and language arts. Math was mentioned far less as a targeted area of reform. This may explain the lack of growth found in math.

CONCLUSIONS AND SUGGESTIONS

This study was an initial exploration into the relationship between data use and student achievement in struggling elementary schools. Below are selected conclusions and suggestions that emerged from this study.

Data Use

The field of education has not necessarily developed a strong and concrete set of best practices around data use. The term itself, data use, was not used consistently across schools in this study; for some, it meant looking at the state testing data once a year, for others, it meant infusing their practice with more reflection and analysis of various forms of data and evidence. All schools included data into their transformation plans, but we do not know enough about the degree to which data really are an integral part of school reform planning and instructional decision making. Having more schools in our sample of case study sites would have provided more insight into the variety of methods used to integrate data into practice. However, because of the timing and the low participation rate among the 13 schools, we were only able to get four schools into our case study sample. What we do know from the four case study schools, however, is that data use is a varied practice (see Data Use Vignettes, Appendix C). Some schools draw on data to make adjustments or decisions more than others, and such inclinations seem to be person-driven or climate-driven (e.g., a champion or team of people on site who really believe in gathering, analyzing, and using data). And we found that at least some teachers and principals at these schools have a growing familiarity and facility with assessment results, which was not always as apparent in the written documents produced by the school (e.g., the transformation plan).

Based primarily on case study data, it is clear that schools need more support in figuring out how to make the data they have access to more useful to them. Principals at the four case study sites were often the conduits, so the degree to which information filtered down to teachers was often dependent on the data-sophistication and commitment of the principal. However, principals may need more support themselves in interpreting and making sense of the data, as indicated by one of our case study principals at Polk. Teachers and principals had many specific critiques of the testing data they receive from the district, and with those, many suggestions for improvements, including: the timing of when they receive data, the format it comes in, and specific areas of training that they request.

Schools' Critiques of the District-Provided Data

Critiques of the data were generally directed at the results they received from the state-testing program. Although not all teachers expressed all of these concerns, the types of concerns raised indicated a growing knowledge, familiarity, and sophistication with assessment results.

- *Change in tests*: Several teachers and some principals lamented the fact that the state changes the test from grade to grade, from ITBS to WASL and back. This

appeared to trouble several teachers who felt that it even further exacerbated the difficulty of tracking individual students' progress over time, and confused interpretation (especially if a student did well on one but not the other). This concern may well be addressed by planned changes in assessment practices proposed by the state (to expand WASL testing into other grades).

- *Lack of diagnostic information*: Some schools mentioned liking the DRA and DWA results because they are more periodic and provide information that teachers feel they can act upon. This may suggest that the district might spend more resources displaying and disseminating reports from these assessments—and other local assessments—to provide schools with the type of data they find more fruitful to their practice.
- *High stakes nature of data*: For the state assessment results, schools have a more difficult time making the data relevant to their practice. Because these data are used primarily for accountability purposes, more work could be done in helping schools understand why it is relevant to their practice, (e.g., do they know how to access or make sense of item analyses on each test?). Are the reports in a format that they find helpful? Schools also feel torn about focusing on other data when WASL is so important for accountability purposes (e.g., AYP). Upper-grade teachers, particularly 4th-grade teachers, feel particular pressure and responsibility for the success or failure of their school in AYP.

Schools' Suggestions to Improve Usefulness of Data

When teachers and principals were asked what might increase or improve data use, teachers across the four schools proposed a variety of ideas for improvement. Most frequently, teachers and principals recommended changes regarding the use and dissemination of district data in the following areas: (a) timeliness of data delivery; (b)increased on-site training and (c) support for assessment and data analysis; and a shift towards individual versus group data reports.

- *Timeliness of data delivery*: Teachers (and principals) at all four schools mentioned that they typically do not get state assessment results early enough in the year to effectively integrate them into practice. Teachers explained that while district data such as DRA or DWA scores arrive within the same academic year, WASL and ITBS data come back "in bits and pieces" after the school year has ended, making them much less useful in terms of adjusting instruction for the same students tested.
- *Individual versus group data report*s: Finally, while they acknowledged that it would be cumbersome and somewhat impractical, a number of teachers stated that they would prefer to receive individual student data to whole class or grade-level data, making it easier for them to track student progress over time and adjust instruction accordingly. The principal at Wilson also suggested that individual student-level reports of the value-added data would be beneficial. While she recognized that it might be important to look at a whole class, she felt that when the data are "lumped" together there is more of a tendency to "point fingers" at teachers

and use data punitively. Furthermore, a group of 4th-grade teachers at Van Buren requested that schools receive students' writing samples from the assessment, along with the grading rubrics or even benchmark samples (examples of highly proficient and not proficient), none of which is currently provided according to teachers. Teachers apparently only get the raw scores from the writing assessment, which many teachers feel, on their own, are not enough to adequately address areas for improvement. Requests to the district for this information were apparently unproductive.

- *Training and support in assessment and data analysis:* As previously mentioned, teachers and principals across the four sites suggested increased training in assessment, data analysis, and interpretation of data as a means of making district data more useful. Many teachers felt that they could particularly benefit from more support in interpreting district data, since they felt that they were not currently using the data to their students' advantage. As an example, the principal at Wilson suggested that use of data would be more prevalent and effective if more time were spent "teaching people how to work with the children to prepare them for tests [and] identify test-taking skills that can be brought into the curriculum," as well as the integration of data use skills and knowledge into content area instruction and curriculum development. Many teachers also recommended that the district provide syntheses and tailored interpretations. For example, one teacher at Truman suggested, "A little one-page summary of 'This is what [the data] looks like, and this is what we need you to do.'"
- *Other kinds of support needed:* Polk serves a high-risk population, and was particularly struggling to meet accountability requirements as well as to integrate data use into practice. The principal felt that he could better help his staff and students to improve if he himself could be provided with more support and training in data analysis and use, but, more urgently, improved support for social services from the district. Without this support, he felt certain that the school would continue to struggle in improving student achievement as outlined in the school transformation plan. Teacher feedback overwhelmingly underscored the staff's collective frustrations with measuring student progress by standardized data alone, without providing adequate support for the social and/or financial problems faced by students at the school. Teachers also disliked being compared with other schools who are not experiencing the same dramatic social issues seen at this school that clearly impact student learning, including transiency, homelessness, drug problems, among many others. As one 4th-grade teacher explained:

> Another piece of support that we don't get is the support that these kids need for social and emotional situations, which impact student learning a great deal. In schools like ours where we've got a high need population and a high poverty index. I'm not making any statement about their ability to learn. I think that they can. They have many more needs than kids who come to this school that are well loved and well fed and well dressed and all that kind of stuff. We don't have support like that. They've taken away our family support worker. They've taken away our counselor. Our nurse is gone away

> most of the time now. We'll really have a nurse one day or half a day next year, something like that. For these kids, in this population, all of these things affect their academic growth …. All that kind of stuff is the support I need. I'm not running to Fred Meyer buying school supplies—(I'm spending time) finding kids clothes … We don't have that kind of support. In my opinion, the Transformation Plan does not address any of those issues at all … We get from the district, 'Teach the whole child.' They only provide the academic aspect. They're expecting us to do the entire rest of that piece, on top of it. For some schools, this is not a problem. Kids come well-fed, lots of parent participation. You can hammer into me all you want about low math scores or low reading scores or low whatever. I absolutely support that. We need to set higher standards and teach to that. We also need help with this other piece. We so critically need help with this other piece. It's just not fair. It's like a lopsided wheel. When one part of the wheel is flat, the whole thing is going to go pluhmp, pluhmp, pluhmp, along. Whereas other kids have a well-rounded wheel. The kids we have are very deserving of a well-rounded wheel. They just, because of life circumstances, don't. We do the best that we can to provide that. I'll tell you, there are only so many hours in a day. There's only so much energy we can devote to it … What it's boiled down to, it's an industrial model being pushed down on a human endeavor. The industrial model is the model of profit and loss, margins. It's being pressed down onto the human endeavor of teaching in human interactions. Education nowadays is so much more than just getting up and teaching them to spell. It's about relationships. It's about human endeavors. It's about interaction.

The climate and social issues evident throughout interview data for Polk were also reflected in its 2002–03 transformation plan. While the primary goals of the plan were clearly focused on district data and improving student achievement on the WASL and ITBS, a concern for creating a "healthy (and) supportive school culture and climate" was also apparent throughout. Beyond a need for academic rigor, the plan states the need for creating a culture of collaboration, caring and respect for cultural diversity, affirmation for diverse opinions in curricular decision making, "staff health (both mental and physical)", infusing multicultural content into curriculum, and understanding that "critical to creating a school culture and climate is for staff and student(s) to do their part in establishing an environment that encourages emotional security."

District Support and Guidance

As a result of the district's efforts, however, schools are well aware of the transformation process and the data they are expected to use (including value-added data). Schools have access to so much data they are often overwhelmed. Moreover, some are not

fully convinced that the data they have access to is beneficial or relevant, especially given the alternating testing program (ITBS/WASL). As mentioned above, schools are torn about focusing on other data sources when WASL makes or breaks their school through AYP. The District could provide more guidance in focusing their attention—either on particular aspects of each test, or on more detailed and disaggregated information on each test that is more tailored to improving instruction (e.g., item analyses or individual growth data). The value-added data, in particular, seems to be a great source of confusion and frustration for many teachers. They do not understand the justification or soundness of comparing results on two very different tests. While the technical explanation might be available, it is not credible or convincing to some. Also, even for those schools who are actively engaged in pre–post testing on CBAs, more support may be needed to guide them in making valid inferences and, importantly, figuring out next steps when results show students need additional help.

The transformation process, according to our site visits, appears to have afforded schools the opportunity to bring teachers together to talk about priorities, something they rarely have a chance to do. For some schools, this time was extremely productive and positive. As mentioned earlier in this report, some schools in this last year may be re-centralizing the planning effort. After the first couple of years of collaborative discussions, some schools have just the principal or the BLT re-write their plan, with some limited feedback from the staff-at-large. This may call into question the degree to which schools are fostering buy-in as they centralize the process.

The data use aspect of transformation planning has been implemented very differently at each site. The transformation planning guidance from the district has changed over time in its focus and requirements for data use. The transformation plan itself has changed in terms of what the district requires (tables, charts, etc.). The role of the coach, the coach herself, and the support provided by the data coach also have changed over time. As a result, the district message about what data use should look like may be unclear to schools. The dominance of and reliance on a table of activities—rather than narrative—in the last round of transformation plans might have contributed to constraining the planning process, by limiting the amount of information about how activities tied together, or the process used to monitor progress. Future plans might place more focus on what structures and processes schools have in place to monitor and evaluate their own progress on the activities and strategies listed.

Relationship between Data Use and Achievement

We cannot make any definitive statements about the relationship between data use and achievement with such a small sample, especially since our richest source of data—the case

studies—only constitutes four schools. We can say, however, a great deal about the individual components of the relationship in these schools: data use and achievement. We know that data use is varied, requires some championing at each site, and needs much more support from the district or outside sources. We also know that there are many ways to look at performance but, in our measures of growth, many of the schools that ranked highly on data use practices also were demonstrating great improvement from pre- to post-transformation periods. Among the 13 schools, there was a lot of overlap in the ratings of data use practice and the ratings of their gains over time. For those schools where their ratings did not overlap, if we had case study data, we were able to hypothesize about what might cause the disconnect between a focus on data use to improve practice and the lack of progress shown. For others who were not in the case study sample, this was not possible. In future studies, the additional case studies would be a crucial element to help illuminate possible reasons for disparities between data use and achievement ratings.

REFERENCES

Choi, K., Seltzer, M., Herman J. L., & Yamashiro. K. (2004). *Children left behind in AYP and non-AYP schools: Using student progress and the distribution of student gains to validate AYP* (CRESST Tech. Rep. No. 637). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Deming, W. E., (1982). *Out of crisis.* Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Heritage, M. & Yeagley, R. (2005). Data use and school improvement: Challenges and prospects. In J. L. Herman & E. H. Haertel, (Eds.), *Uses and misuses of data for educational accountability and improvement. National Society for the Study of Education. Yearbook of the National Society for the Study of Education Vol. 104* (Issue 2, pp. 320–339). Chicago: Blackwell Publishing

Herman, J. L. (2005, April). *Developing teachers' expertise in L2 academic language: A functional grammar approach*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Montréal, Quebec, Canada.

Herman, J. L. & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation* (CRESST Tech. Rep. No. 535). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2005, April). The nature and impact of teachers' formative assessment practices. In J. L. Herman (Chair), *Building science assessment systems that serve accountability and student learning: The CAESL model.* Symposium conducted at the annual meeting of the American Educational Research Association (AERA), Montréal, Quebec, Canada.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Seltzer, M. & Choi, K. (2002). Model checking and sensitivity analysis for multilevel models. In N. Duan & S. P. Reise (Eds.), *Multilevel modeling: Methodological advances, issues, and applications.* Mahwah, NJ: Lawrence Earlbaum Associates.

Seltzer, M., Choi, K. & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Implications for conducting analyses that help illuminate the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*(3), 263–286.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Currency Doubleday.

The White House (2001). *Transforming the Federal role in education so that no child is left behind*. Retrieved at www.whitehouse.gov/news/reports/no-child-left-behind.html

Tyler, R. W. (1949). *Basic principles of curriculum and instruction.* Chicago: University of Chicago Press.

APPENDIX A:

LIST OF SCHOOLS

School	Total enrollment	Grade level
Total	46,416	K–12
Carter	370	K–5
Fillmore	169	K–5
Harding	258	K–5
Hoover	533	Pre K–5
Jefferson	167	K–5
Kennedy	430	K–5
Lincoln	519	K–5
Pierce	232	K–5
Polk	134	K–5
Truman	208	K–5
Tyler	253	K–5
Van Buren	299	Pre K–5
Wilson	295	K–5

Note. K = kindergarten.

APPENDIX B:

TRANSFORMATION PLAN REVIEW RUBRIC

Area	Factor	Rating
Evidence / Indicators Used	Breadth/Coverage	1 = state/district assessments (WASL, ITBS, DRA, DWA) 2 = use of CBAs, VAT, portfolios, etc. 3 = comprehensive indicator system (e.g., system includes surveys, portfolios, tracking/monitoring activities, and other types of indicators beyond standardized tests)
	Depth	1 = district-reported trends (comparison to selves & standard) 2 = beyond trends (desegregations or comparisons with others on subskills, subtests, gains on VAT, etc.) 3 = formative, diagnostic assessments delving into problems
	VAT	1 = no mention of VAT 2 = mention VAT in terms of trends 3 = extensive mention/use of VAT in their analysis
	Technical Sophistication	1 = inaccurate, inappropriate 2 = accurate, appropriate 3 = synthesis/triangulation/attention to reliability
	Improvement O/T	1 = no improvement 2 = some improvement 3 = extreme improvement
Identification of Needs / Objectives / Problem	Breadth/Depth	1 = coverage of district achievement goals, including disproportionality 2 = specific benchmark *objectives* to address district goals or additional objectives beyond achievement goals 3 = comprehensive and specific set of objectives and goals
	Improvement O/T	1 = no improvement 2 = some improvement 3 = extreme improvement
Identification of Solution Strategies	Specificity (e.g., benchmarks)	1 = no sub-strategies or action steps identified 2 = some sub-strategies or action steps identified 3 = system of measurable, specific sub-strategies/action steps
	Theory-based / Data-driven / Research-based	1 = seemingly random 2 = linked to problems or needs identified thru data 3 = research-based, with strong rationale
	Improvement O/T	1 = no improvement 2 = some improvement 3 = extreme improvement

(*table continues*)

Area	Factor	Rating
Analysis of Progress	Formative/periodic	1 = no indication that analysis is ongoing and periodic (annual) 2 = periodic review of analysis (more than annual) 3 = periodic, ongoing, built into decision-making progress
	Improvement O/T	1 = no improvement 2 = some improvement 3 = extreme improvement
Inclusive Decision-Making	Included throughout Process	1 = no sign of inclusion 2 = signs of typical involvement—PTA, BLT—but limited 3 = intensive inclusion throughout process
	Improvement O/T	1 = no improvement 2 = some improvement 3 = extreme improvement

Note. WASL = Washington Assessment of Student Learning, ITBS = Iowa Test of Basic Skills, DRA = Developmental Reading Assessment, DWA = Direct Writing Assessment, CBA = curriculum-based assessment, VAT = Value-added Test Scores, O/T = over time, PTA = Parent Teacher Association, BLT = Building Leadership Team.

APPENDIX C:

VIGNETTE DESCRIPTIONS OF DATA USE PRACTICES AND MONITORING FOUR CASE STUDY SITES

Van Buren

The principal at Van Buren devotes a great deal of time to processing data to make it more understandable and applicable to the teaching staff. She explained that, "when I get the data, when it arrives here, I'm always trying to figure out some way to digest it so that [the teachers] can digest it ... looking for a way to make it meaningful and digestible to [them]." Along this line, throughout the interview, she described the various types of data analysis she does to provide useful information to the teachers. First, she does a data summary, where she uses a technique she learned from a professor at the University of Washington, to denote what areas are "hot" or "cold," meaning areas for that are going well and areas in need of improvement, respectively. She then shares this information with the teachers to help them understand how they are doing. Second, she compiles lists of individual students in a classroom, showing each student's test results (applicable data varies), to help teachers see a student's strengths and weaknesses, identify who is at standard, or close to reaching standard, and then help them to achieve this. This data is used to inform teacher practice, or if a student needs to be in a reading intervention program, for example. Furthermore, the principal mentioned that she processes the staff climate survey data and rank orders the items to show their strengths and weaknesses. She also compares the school's results to that of the district-wide results, and shares this with the teachers.

When developing the school's transformation plan, members of the BLT recounted that the school looks at its current status, operation, and student results, along with goals and standards, and then begins the process of writing a transformation plan. Once she began working at the school, the principal used the *Johnson City Model* of organization development to help shape the plan. With that model, they look at what they know from their data, what teachers believe, and then identify what is positive, and what needed to be improved upon. A collaborative model was used, with principal leadership, BLT involvement, and decisions by staff consensus. An example of how data were used to inform the development of the transformation plan is that, as members of the BLT described, they had initially planned to focus on writing, but after looking at data showing that reading was an area in need of improvement, the focus was shifted. Additionally, the primary-grade teachers explained that when revising the plan they looked at student survey data revealing that students felt bullied, therefore, strategies to address this issue were added to the plan.

While reading has been the focus of the transformation plan goals, the school will be working on writing next. According to the principal, data were used to show that, specifically, informational text reading and math measurement were areas to work on. Further explaining the transformation plan goals, the principal said that one of the goals was to increase the time spent on reading at the school, and they worked on using the Accelerated Readers program and NUA as strategies to reach this goal. Another goal was to increase the collaborative time for teachers to meet to improve their practices, and NUA strategies were often discussed during this time.

WASL scores are used to inform teacher practice when, for example, the data shows that there is an area where students are not testing well, like math measurement. As previously discussed, the 4th-grade teachers revised their teaching strategies in this area to help raise the student's scores. Expository writing also was an area that needed improvement as well, and teachers expressed that they felt frustrated that they do not get the writing test data back to be able to see the specific areas that need to be focused on. "That's one of our complaints," a 4th-grade teacher explained, "the test should be a teaching tool for us to help us guide our teaching on where our kids are falling down. We've never seen the results of the writing in particular." Furthermore, a teacher stated, "I think that it's pretty clear that we use a ton of assessment material to guide our teaching and also just your classroom assessment."

According to the principal, monitoring the effectiveness of the of the transformation plan goals at Van Buren is done by asking questions such as, "Did we do it? Are we doing it? Are we seeing results with student data?" Seeing progress through student work based on the goals is also an indicator of progress. For example, the principal cited that a goal in the 4th grade was to help students improve writing strategies, and towards the end of the year, the teachers brought portfolios of student work to the principal to demonstrate that "100% of their kids, when given a prompt or given a problem, could do this…" Classroom data and other assessments are also used to measure progress.

Additionally, members of the BLT shared several ways by which the effectiveness of the transformation plan is monitored that include: students transitioning to the next grade level as better readers and writers; data reports that the principal brings to meetings showing positive results; and an increase in the number of students participating in the school-wide home reading program. A teacher also jokingly added, "when you see your name in the paper in that list of schools, you don't have to black out your windows on the way to school" (referring to the newspaper's report of standardized test results). Finally, a teacher on the BLT stated, "I think a big impact has been the awareness of all grade teachers knowing what other grades are doing and being willing to give up things for the greater good. They realize

how important things are in the upper grades." The move to a full day kindergarten program at the school illustrates this point, as the teachers have placed great importance in starting the process in the early grades. Furthermore, BLT members shared that having the transformation plan, "gives us direction." A teacher then added, "Because it was formed by the whole staff, we all kind of bought into what the goals are." While another teacher chimed in, "I think you see it more as a process. It begins in kindergarten. Each grade level is not just isolated. You are all working toward this goal of students performing at a certain level on a test or improving reading scores. So everyone has a part in that." As a result, these teachers shared, the plan has opened up conversations across grade levels about the types of things they are doing in their classrooms.

Information from teacher's classroom-based assignments and assessments, observations, and "professional judgment" are some of the methods by which teachers identify low-performing students. Additionally, data from formal assessments such as the previous year's WASL or ITBS results, IRIs, DRAs, and CBAs are used. When teachers have real concerns, they go to the student intervention team (SIT) to discuss strategies to help the student. With these various types of information, the principal suggested that "teachers are 80% right" in identifying students who need extra assistance by looking at data and through direct observation. Another way teachers gather information on their students is through portfolios (interviewee did not specify what type of information is included). However, the 4th-grade teachers said that they do not tend to look at the portfolio ahead of time, but if they have a concern they will look back to see what their previous scores were, for example. A kindergarten teacher shared that the reading intervention programs are used as a way to know a student needs extra help.

Data is aggregated at the student level through the principal's hot and cold lists. Different pieces of information are brought together on each student (e.g., DWA, ITBS, and information on their language status or involvement in an intervention program) to help teachers know where their weakness are and how they can help them to achieve standard if they know, for example, that they were previously very close to doing so. Using samples of student's work throughout the school year is also another tool for monitoring growth. A first-grade teacher shared an experience that illustrated the usefulness of assessing students on similar tests from year to year, as when she had a concern about a student she did a DRA with him. The score was not particularly great, but then she looked back to his previous years score on the DRA and was able to judge that, in fact, the student was progressing.

Wilson

Many types of data are used to inform the development of the transformation plan at Wilson, according to the principal. At the beginning and end of each year, she reviews the testing and assessment data, such as the WASL and DRA, with the teachers. Staff surveys about the environment of the school, discipline records, and student surveys are used as well. The data are reviewed by principal presentation or small group discussion, depending on what they are focusing on. When results are discussed at the beginning of the school year, it is sometimes necessary to make minor revisions to the transformation plan goals, as the plan is written in the previous school year before they have seen the data. Furthermore, teachers use their own data that they collect throughout the year to map out what they want to do in the future. Initially when the transformation plan was being developed, the principal, BLT, and staff were all involved, to a degree, and a significant amount of time was spent in this process. It is a highly collaborative process. Also, when reviewing data and identifying areas for improvement, the principal said that they try to find ways to improve their practice starting with the early grades, so that by the time the testing start, the students are better prepared. Still, she explained that teachers are doing test prep with the students, and as a result, they can identify areas where the students had trouble understanding and then teachers can "fill in those holes."

According to the principal, data review occurs throughout the school year in various ways. Teachers frequently discuss how the students are progressing and how assessment results are reflected across grade levels (e.g., how teachers in one grade can prepare students for the following grade). The primary-grade teachers shared that they interpret the CBA data themselves, and then meet with their teaching teams to discuss the results. In addition, a few teachers have worked with a district consultant, brought in by the principal, to look at data. Specifically, these teachers tracked individual students in math to see where they started and finished, what seemed to work, and what did not. In addition, in looking at math data, teachers at the school discovered that measurement was a real weakness, and then concluded that this was because these concepts were not taught until after the test was taken. Consequently, they decided to start teaching measurement once a week from the beginning of the school year. The principal added, "so it's that way that we look at [data] in the course of the year, and think about it in the course of the year, and it comes up again as we're doing our planning."

Furthermore, the principal described that with the ITBS and WASL item analysis provided by the district, she and the teachers looked at the test items, especially those that the students had not done as well. With this data, as well as information such as what is categorized for each of the scoring criteria, the principal and teachers tried to determine how

they could improve the program. To illustrate, the principal explained, "we looked at [the test items] and tried to figure out, what is it in this item that maybe our children haven't learned from us that we could improve or increase? What kind of thinking were we expecting?"

The principal shared that she had noticed that there were discrepancies between results from district assessments, and the classroom data collected by teachers. This led to a discussion with the teachers and then she decided to look further into the reasons why this was happening. In particular, she looked at student data from the DRA and ITBS and was interested in looking at individual students' growth over the years. To illustrate, the principal remarked, "I always goes back and look at the fifth-graders writing, and how they did in fourth-grade, and how they did in third grade." She noted that she had recently discovered that although the 5th-grade students' results may not have been great, that, in fact, there had been an increase in the number of students passing from one year to the next. Then, there was an attempt to figure why scores were incongruent across grade level and why they might have been this way. For example, the DRA is administered one-on-one (teacher works individually with each student), while students work independently on the ITBS. Noting another discrepancy in students' reading scores on ITBS, through the work with the Value-Added coach, they uncovered that it seems that students have difficulty following and reading the directions. Therefore, time was spent in school on developing this skill.

Additionally, a teacher stated that using information on student performance is "the most powerful way to change and direct curriculum … [by looking at the] performance of students and reflecting about it—thinking about it, questioning it, questioning your practices, questioning your strategies and trying what is sort of suggestion by your reflections. So, somehow I feel much more confidant [*sic*] working on performance, rather than data and the numbers."

When asked about how progress towards the transformation plan goals is monitored at Wilson, the principal stated that, "I figure we judge it based on how the children are doing every day at school, and what the climate of the school is, and what's happening with the test scores. How we all subjectively feel about how we're doing here." She continued by saying that assessment results and teachers' sense of how they can further "stretch and push" their students, and then adjust their teaching, are also factors. Several teachers shared that evidence of progress can often be seen by walking around the classrooms and seeing that the strategies are being used. Moreover, the principal responded to the question of how the effectiveness of the transformation plan is determined, by describing that she and the teachers will see improvements in the students' reading and the level and complexity of what they are able to do in the classroom. While test scores and teacher-based assessments are used, the

types of books students are reading and the depth of their discussions about the books are noted as well. Finally, the principal said that teachers would share that they tried something new or different and saw a positive improvement for the students. For example, she said that teachers will come to her office to show her their student's work and, "that, to me, is a piece of data," she concluded.

Teachers cited several ways in which the transformation plan has had a school-wide impact. For example, a teacher stated that it guides the types of professional development teachers are encouraged to participate in, such as GLAD or NUA, or the kinds of staff development they do at the school. The plan also helps to focus improvement strategies, as there are so many things that can be done, if it is not written in the plan it can be cut, a 3rd/4th-grade teacher described. As such, it helps to prioritize their goals and to be focused on what they are trying to achieve.

Polk

Data use at Polk appears to be focused mostly on identifying individual student needs and areas for improvement. The principal has instituted a process across grade levels for utilizing student-level data to inform instruction and address low student test scores. Grade-level teachers college standardized test data for each student. If, according to the data, a student is performing below grade level, teachers meet with a team of colleagues, or Student Intervention Team (SIT), to formulate a plan to help the student improve.

Teachers are also required to use student data to develop Student Acceleration Plans (SAPs) for all students, which provide academic goals for each student on a yearly basis. The principal explained that teachers must, "identify the data and… identify what strategy [they] are going to use to accelerate that student. Then [they provide] other basic information about the student, whether there be behavior issues, home issues, attendance issues, or health issues." In each SAP, teachers must provide individual academic goals and the means or strategies to achieve them. Teachers meet regularly, both formally and informally, to discuss student goals, issues, and improvement. While some teachers mentioned discussing strategies with a student's teacher(s) from a previous year, most described meeting with other grade-level teachers to discuss student data. One first-grade teacher commented, "One meeting might be grade level meetings. So we have a chance to talk with our grade level counterpart and share ideas and check what they are doing in their room, and share data." While the SIT and SAP processes are essentially data-focused, teacher and principal feedback indicates that their approach is more holistic than entirely data-driven (i.e., consideration of social and personal issues appear to be essential to these plans).

Another theme that emerged across interviews regarding data use was that of collaboration. As described above, teachers meet frequently, both formally and informally, to discuss individual student data and progress. Additionally, the staff meets school-wide at least twice a year as a group to "review the data from the standardized tests across the school," according to one 2nd-grade teacher. In these meetings, staff and principal share classroom data, compare grade-, school-, and district-level data, and, as the same teacher explained, "determine where we need to show growth—[determine] what we need to work on." Teachers also use this collaborative time to discuss student growth as it relates to the goals of the Transformation Plan.

Truman

With regard to using data to identify areas in need of improvement at Truman, the school's reading specialists stated, "The way it's set up, you can visually see the areas that we need to work on and the areas we're being successful in." She then correlated her statement with Truman's test scores and stated, "It's not where we want to be, but we have looked at data and determined where our children need to be and what we need to do to make this happen." Some of the teachers stated that they now test more and look at data more without providing any other details. Another teacher reported that the Standardized Testing and Reporting (STAR) data do not match very well with the district's DWA and DRA because the data provide two different kinds of results. The upper-grades elementary teachers reported that they look at the data, hold meetings, and collaborate on what the scores mean, where the weaknesses lie, and how best to approach instruction and strategies to improve test scores. One teacher stated:

> For the OSPI grant, we take the math portion and we divide it by the strands: geometry, fractions, statistics, number sense, and probability. We look at where the kids are at. We might have meetings where we talk with the staff. We haven't gone real deep into each strand yet, but that's part of the honing in on the math part. In a roundabout way, we take the WASL scores from the previous year, and we use that to see which strands we need to be focused on. With that, we tutor and we develop tests for all 4th grade students who are taking the WASL. We took the weakest strands and I developed the curriculum.

Monitoring and accountability is executed in different ways, depending on the individual's role at Truman. First, the principal reported that she sets outcomes that she expects to see. She provides a range of examples as to how she monitors and accounts for different components surrounding curriculum and instruction. She stated:

> If it's around discipline, you see that the behavior matches what you expect the outcome to be. If it's around literacy, you see the improved academic performance around kids. If it's around mathematics, you see children behaving in a mathematical way … If it's around professional staff development, you see teachers using better teaching practices in the classroom. If it's around adult communications, you see adults responding in a different way. You see the action that you set out for initially occurring in the classroom or in the school community.

Furthermore, the reading specialist reported that she measures progress by looking at where students are at the beginning and where they currently are performing. For example, students are tested at the beginning of the year using STAR, and if results show they are in need of improvement, additional testing and assistance is provided.

Likewise, the upper elementary teachers reported that they monitor progress on not only assessment data gains, but also on how well their students are engaged in their work. One teacher stated, "When they're excited and talking about it and they're using the vocabulary, you can see the little light bulb go on; 'Gotcha, I got it.'" Improved student behavior has also been cited as a means to monitor progress. Additionally, the upper elementary teachers are accountable to each other. As one teacher stated, "And the teams meet together to look at what fifth-grade teachers talking to the fourth-grade teachers about what they're going to need to know in fifth-grade and what kids are coming with, and trying to fill in the gaps on that. Using each other to see, 'this isn't being taught, and this needs to be taught before they come into your grade. What is it that I as a fourth-grade teacher need to be teaching so the kids can come into your fifth-grade class and be ready to start?'" Another teacher stated, "That's the other good thing about those lesson studies, because you can watch [other teachers] do a fifth-grade lesson with the third graders. We will go in and watch the same lesson, but on the third-grade level, so you can compare and say, 'As a fourth-grade teacher I'm in the middle of that.' You can see the different ability levels in a grade, so you can [figure out] where they're coming from."

The upper elementary teachers also reported that they are not only accountable to their students and to each other, but also to their principal. That is, their principal has expectations of teachers related to students' learning and experiences in class. As one teacher stated, "We have expectations that were placed upon us by our principal. We hold good to it, and we value that. We realize it's only going to make us better. We expect children to be learning, she expects for us to be learning to help them learn."

Another method by which progress is monitored, according to the principal, is by conducting classroom observations with the assistance of both her head teacher and the

reading teacher. The principal reported that the upper elementary teachers' progress is primarily monitored by peer review in the form of peer classroom observations. Parent events are held on a monthly basis, where teachers tell what they are seeing, what they do not see, and what they would like to see. Finally, the BLT monitors direct progress based on what is written in Truman's transformation plan. The BLT will raise such questions as "Have we done it? What else do we need to do?" Similar questions are posed to the principal by her supervisor, and through this process, the principal visits the transformation plan quite frequently. Some of the questions asked by the supervisor are as follows: "Tell us where you are in your transformation plan. You guys said you were going to do this. Are you guys doing a template that looks like this?"

The identification of low-performing students begins prior to the beginning of the school year. That is, prior to school opening for the year, the school improvement team conducts home visits and gathers data. The teachers convene and share information on students' strengths and weaknesses, and consider the data provided from the home visits. Based on this information, students are labeled as high-impact, medium-impact, or low-impact. High-impact students are those students who need the most support and an intervention plan is developed for each of these students. A support person, who may be the head teacher, reading teacher, principal, an intern from the local university, or a counselor, is assigned to the student to provide individual support. Additionally, the STAR assessment and a placement test are administered at the beginning of the school year, and the data are analyzed immediately. These data provide information for grouping students. Anecdotal notes and teacher observation are used throughout the year to monitor student progress.