U.S. DEPARTMENT OF EDUCATION

Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports

Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports

U.S. Department of Education
Office of Planning, Evaluation and Policy Development

Prepared by:
Barbara Means
Eva Chen
Angela DeBarger
Christine Padilla

SRI International

2011

This report was prepared for the U.S. Department of Education under Contract number ED-01-CO-0040/0002 with SRI International. Bernadette Adams Yates served as the project manager. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education is intended or should be inferred.

February 2011

Contents

Exhibits

Figures and Tables

Executive Summary

The national Study of Education Data Systems and Decision Making documented the availability and features of education data systems and the prevalence and nature of data-informed decision making in districts and schools. That study, like past research, found that teachers' likelihood of using data in decision making is affected by how confident they feel about their knowledge and skills in data analysis and data interpretation (U.S. Department of Education 2008). Unfortunately, teacher training programs generally have not addressed data skills and data-informed decision-making processes. Understanding the nature of teachers' proficiencies and difficulties in data use is important for providing appropriate training and support to teachers, because they are expected to use student data as a basis for improving the effectiveness of their practice.

This report describes an exploratory substudy on teachers' thinking about data conducted in conjunction with the larger Study of Education Data Systems and Decision Making data collection and the implications of the substudy findings for teacher preparation and support. Teachers' thinking about student data was investigated by administering interviews using a set of hypothetical education scenarios accompanied by standard data displays and questions to teachers in schools within case study districts selected as exemplars of active data use through an expert nomination process.[1] Data scenarios were administered to both individual teachers and small groups of educators who typically work together. Conducting both individual and group interviews provided information about how teachers reason independently about data as well as about how they build on each other's understanding when they explore data in small groups.

Part I of this report describes the responses that 50 individual teachers and 72 small groups gave to the data scenario questions. The teachers interviewed are not a nationally representative sample, but they do provide a detailed initial look at how these particular teachers think about student data in schools that were thought to be ahead of most schools in the nation with respect to data use. This detailed description of teacher thinking can help inform those responsible for training teachers in data-driven decision making about the kinds of difficulties and misconceptions teachers are likely to encounter. Part II of this report provides material that can be used in training teachers on the use of data to guide instruction. It contains the seven data scenarios used in the exploratory substudy, along with guidance on how a professional development provider can use the scenarios as part of teacher training or teacher learning community activities and the particular points that should be looked for in teachers' responses to each scenario.

[1] The case study districts represented a diverse group in terms of the number of students served (5,559 to 164,000), the percentage of minority students (17 percent to 83 percent), and the percentage of students who qualify for free or reduced-price lunches (13 percent to 64 percent), as well as in terms of urbanicity (urban, suburban, and rural) and regional location. Within each district, three schools were identified: one school that the district considered advanced in its data use practices, one school that was typical of the district in its level of data use, and one school that was emerging as a strong data user.

Working with external data system and measurement experts, the research team identified five skill areas that cover the different aspects of data use that the experts thought teachers need to master if they are to use student data to improve instruction. Teachers need to be able to do the following:

- Find the relevant pieces of data in the data system or display available to them (*data location*)
- Understand what the data signify (*data comprehension*)
- Figure out what the data mean (*data interpretation*)
- Select an instructional approach that addresses the situation identified through the data (*instructional decision making*)
- Frame instructionally relevant questions that can be addressed by the data in the system (*question posing*)

Data scenario interviews were designed to tap into these five components of data literacy and use. Data location skills are essential for identifying data that will be used to inform teachers' decisions about students. Data comprehension skills, such as understanding the meaning of a particular type of data display (e.g., a histogram) or representing data in different ways, are necessary for figuring out what data say. Data interpretation skills are required for teachers to make meaning of the data. More sophisticated data interpretation skills, such as understanding the concept of measurement error and score reliability, are particularly important for administrators who need to make more high stakes decisions about students and teachers based on test performance. As data systems become more readily available to teachers, the ability to pose questions that generate useful data will become increasingly important.

Interview Responses

During the 2007–08 school year, the research team collected data scenario responses from 52 individual teachers and 70 small groups of school staff from 21 elementary schools and 14 middle schools across 13 school districts located in 12 different states.[2] The data scenario interviews were administered as part of site visits to schools clustered within districts participating in the larger national Study of Education Data Systems and Decision Making. The case study districts were purposefully selected from a set of districts nominated by data-driven decision-making experts on the basis of their active use of student data to inform instruction. The case study districts are not a nationally representative sample. They were selected to represent best cases—sites where student data systems are available to teachers and teachers are encouraged and supported in the use of student data for decision making. These cases permitted

[2] Groups typically consisted of three teachers or two teachers and a school administrator or specialist. A total of 180 teachers and 35 administrators/specialists participated in group interviews.

the research team to collect information from teachers who have experience using student data and data systems.

Interview responses were transcribed and analyzed using a standard coding scheme. The analysis identified strengths and weaknesses related to each component of data literacy.

Data Location

- Teachers in case study schools generally were adept at finding information shown explicitly in a table or graph.

Data Comprehension

- Teachers in case study schools sometimes had difficulty responding to questions that required manipulating and comparing numbers in a complex data display (e.g., computing two percentages and comparing them).
- Some case study teachers' verbal descriptions of data suggested that they failed to distinguish a histogram from a bar graph or to consider the difference between cross-sectional and longitudinal data sets.

Data Interpretation

- Many case study teachers acknowledged that sample size affects the strength of the generalization that can be made from a data set and suggested that any individual student assessment administration may be affected by ephemeral factors (such as a student's illness).
- Case study teachers were more likely to examine score distributions and to think about the potential effect of extremely high or low scores on a group average when shown individual students' scores on a class roster than when looking at tables or graphs showing averages for a grade, school, or district. An implication of this finding is that teachers will need more support when they are expected to make sense of summaries of larger data sets as part of a grade-level, school, or district improvement team.
- Case study teachers' comments showed a limited understanding of such concepts as test validity, score reliability, and measurement error. Without understanding these concepts, teachers are susceptible to invalid inferences, such as assuming that any student who has scored above the proficiency cutoff on a benchmark test (even if just above the cutoff) will attain proficiency on the state accountability test.

Data Use for Instructional Decision Making

- Many case study teachers expressed a desire to see assessment results at the level of subscales (groups of test items) related to specific standards and at the level of individual items in order to tailor their instruction. After years of increased emphasis on accountability, these teachers appeared quite sensitive to the fact that students will do better on a test if they have received instruction on the covered content and have had their learning assessed in the same way (e.g., same item format) in the past.

- Many case study teachers talked about differentiating instruction on the basis of student assessment results. Teachers described grouping strategies, increased instructional time for individual students on topics they are weak on, and alternative instructional approaches.

Question Posing

- In order to use an electronic data system to identify areas for improvement, educators must be able to frame questions that can be addressed by the data in the system. Many case study teachers struggled when trying to pose questions relevant to improving achievement that could be investigated using the data in a typical electronic system. They were more likely to frame questions around student demographic variables (e.g., “Did girls have higher reading achievement scores than boys?”) than around school variables (e.g., “Do student achievement scores vary for different teachers?”).

This exploratory study also revealed that case study school staff working in small groups were more likely than individual teachers to seek clarification of the scenario questions and to catch errors in the information they were looking at or in the computations made. The small groups had a significantly higher probability of giving a correct answer to 5 of the 17 data scenario items on the data literacy scale developed from selected interview items. [3] Teachers in one-on-one interviews had more accurate responses on none of the 17 items (there were no significant differences on 12 items). Teachers responding to the data scenarios in small groups also displayed more indications of enjoying the data scenario interviews, suggesting that school routines for collaborative work with data may foster greater interest in data use among teachers. This finding, coupled with prior research showing that working with data tends to increase the amount of collaboration among teachers (Chen, Heritage, and Lee 2005; Wayman and Stringfield 2006), suggests that working with data and collaborating on instruction may be mutually reinforcing school improvement practices.

[3] The research team developed the data literacy scale from selected items of the data scenarios designed to measure teacher skills in data location, data comprehension, data interpretation, question posing, and data use for instructional decision making. Scale scores were obtained by averaging the item scores.

Implications for Practice

A better understanding of teachers' strengths and weaknesses in understanding data can inform the design of more effective teacher training and professional development. As student populations become more diverse, teachers face the challenge of providing differentiated instruction to students with a wide range of knowledge and skill levels. By improving skills related to collecting, analyzing, and interpreting student assessment data, teachers will be potentially better equipped to adjust their instruction to accommodate the needs of individual students. As teachers have access to more student data, they need to learn to interpret the data themselves to adjust instruction in a timely manner. The skills that are the focus of this study are those experts judge as important if teachers are to understand what their students know, how students perform individually and as a group, what areas of their instruction need improvement, and how to group students and apply tailored strategies. The findings in this study can help inform the design of pre- and in-service teacher training programs on data-informed decision-making processes.

Findings from this exploratory research can be used by schools and districts forming professional learning communities to facilitate teachers' dialogues about data use. A recent What Works Clearinghouse practice guide (Hamilton et al. 2009) suggests that such learning communities can foster teachers' skill in using data to inform instructional decisions.

The data scenarios used to elicit teachers' thinking about data in the substudy can be used by those who facilitate teacher learning in this area. By responding to questions in the data scenarios, teachers and administrators have the opportunity to deepen their understanding of skills and concepts essential to assessment and data analysis. Such professional development activities can be a supplement to, but are no replacement for, teachers' work with the data from their own classrooms, grade, and school. Case study work reported elsewhere (U.S. Department of Education 2010) suggests that teachers' data skills are best developed through ongoing opportunities to examine their own students' data and draw inferences for practice with the support of colleagues and instructional coaches.

Implications for Future Research

Given the importance that federal education policy places on teachers' data literacy and ability to draw instructional implications from data, additional research and development in this area are warranted. The rationale for this exploratory study came from national survey data showing that both district administrators and teachers themselves express reservations about teachers' ability to make sense of student data reports provided by electronic systems. The national data were collected from districts during the 2007–08 school year and from teachers during the 2006–07 school year. As policymakers continue to emphasize data use to support education reforms, it will be important to track progress made in teachers' data literacy and ability to use data. An important next step is validation of assessments of teachers' data skills against measures of actual teacher practice with data.

Part I: An Exploratory Study of Teachers' Data Literacy Skills

1. Introduction and Approach

The *Elementary and Secondary Education Act of 2001* (*ESEA*) and associated guidance promote the use of data to guide decisions about instruction not just at the district level but also at the level of individual schools and teachers' classrooms. Many districts began upgrading their student data systems in response to *ESEA* accountability provisions and requirements for achievement data reporting by student subgroup. Policymakers urged districts to engage in "data-driven decision making" as a complement to "research-based practice" (Consortium for School Networking 2004; U.S. Department of Education 2004). Proponents of data-driven decision making call on educators to adopt a continuous-improvement perspective, with an emphasis on goal setting, measurement, and feedback loops so that they can reflect on their programs and processes, relate them to student outcomes, and make refinements suggested by the outcome data (Schmoker 1996).[4] The emphasis on data systems and data use for educational improvement was further underscored when the education funds to be distributed to states under the *American Recovery and Reinvestment Act of 2009* were made contingent on assurances about the use of student data systems.

The examination of data is not an end in itself but rather a means to improve decisions about instructional programs, student placement, and instructional methods. Once data have been analyzed to reveal areas in which the instruction provided is not generating desired student outcomes or to identify specific students who have not attained the expected level of proficiency, educators need to reflect on the aspects of their practice that may contribute to less-than-desired outcomes and to generate ideas for how they could change their practice in ways that would produce better student outcomes.

The rationale for this exploratory study came from earlier survey data showing that both district administrators and teachers themselves express reservations about teachers' ability to make sense of student data reports provided by electronic systems. For example, the majority of teachers with access to a student data system in 2007 reported that they could benefit from further professional development on a variety of topics related to data use.[5] More than half of the surveyed teachers said that they needed additional professional development on how to adjust their instructional content and approach based on data; 48 percent reported needing professional development on the proper interpretation of test scores; and 38 percent indicated a need for

---

4 Some researchers prefer the term *data-informed decision making* in recognition of the fact that few decisions are based wholly on quantitative data. This report uses the term *data-driven decision making* because of its prevalence; no implication that data should be the sole determinant of actions is intended.

5 The district survey sample of 1,039 districts was nationally representative with respect to poverty status, student enrollment, and location (urban or rural). The 1,799 teachers surveyed in spring 2007 taught core academic subjects within 865 schools nested within the district sample.

training on how to formulate questions that they could address with data (U.S. Department of Education 2008).

In the past, teacher training generally did not include data analysis skills or data-driven decision-making processes (Choppin 2002). Without data skills, teachers are ill prepared to use data effectively to provide instruction that matches students' needs. Moreover, the measurement issues affecting the interpretation of assessment data—and certainly the comparison of data across years, schools, or different student subgroups—are complicated. Data use for gaming the system (Diamond and Cooper 2007) and data misinterpretation are real concerns (Confrey and Makar 2005). For this reason, districts and schools are devoting increasing amounts of professional development time to the topic of data-driven decision making (U.S. Department of Education 2008). Many argue that the practice of bringing teachers together to examine data on their students and relate those data to their practices is a valuable form of professional development in its own right (Feldman and Tung 2001). Some districts have used Enhancing Education Through Technology (EETT) professional development funds to underwrite these activities. In addition, some districts that have been active in this area provide data coaches or other means for accessing technical expertise for school teams engaged in looking at data (U.S. Department of Education 2010).

Understanding the nature of teachers' proficiencies and difficulties in data literacy is an important consideration in designing supports for data-driven decision making and sheds light on gaps in teacher education and professional development programs.

Purpose of the Report

The national Study of Education Data Systems and Decision Making, sponsored by the U.S. Department of Education's Office of Planning, Evaluation and Policy Development, documented the availability of education data systems, their characteristics, and the prevalence and nature of data-driven decision making in districts and schools (U.S. Department of Education 2008, 2009, 2010). The study examined both the implementation of student data systems and the broader set of practices involving the use of data to improve instruction, regardless of whether or not the data are stored in and accessed through an electronic system.

This report describes an exploratory substudy of teachers' responses to a set of scenarios involving hypothetical student data. The scenarios were designed to probe teachers' understanding of the kinds of data available to support their instructional decisions. Teachers participating in the substudy teach within case study districts nominated as exemplary users of student data, as described in the study final report (U.S. Department of Education 2010). These teachers have experienced more support and professional development for data-driven decision making than U.S. teachers as a whole (U.S. Department of Education 2010). Aspects of understanding data displays and data interpretation issues that continue to challenge this group of teachers are strong candidates for focus in teacher preparation and professional development programs generally. Part II of this report provides the data scenarios used in the substudy along with guidance on how they can be used as part of professional development on data literacy and data-driven decision making.

Prior Research on Data Use and Decision Making

Every week, school leaders and teachers make hundreds if not thousands of judgments affecting the instruction that students receive. Whether or not they use data to support their decision making, school staff decide what and how to teach every student in every class.

Cognitive research has highlighted a number of issues in the development of understanding in the areas of statistics and data representations. Bright and Friel (1998), for example, found that when students first start working with graphs, they have difficulty moving back and forth between raw data for individuals and the group data represented in the graph. Curcio (1989) described a developmental progression for graph comprehension, starting with simple reading of graphs and then advancing to the ability to identify mathematical relationships shown in graphs and finally to being able to draw inferences from graphed data. Konold (2002) reported that although people do not have difficulty understanding the idea of covariation, they do have difficulty relating this concept to displays such as scatterplots. Noss, Pozzi, and Hoyles (1999) studied use of data displays by practicing nurses and found that even though the nurses knew that blood pressure increases with age from their own experience and could use software to generate scatterplots of data on individuals' age and blood pressure, they were not able to "see" the relationship between the two variables in a scatterplot of these data.

Psychological research on decision making has identified cognitive processes that lead to biases in decision making, particularly when probabilities are involved—as they necessarily are when trying to anticipate future events or behaviors. A classic review by Tversky and Kahneman (1982) highlighted three biases: (1) representativeness bias, (2) availability bias, and (3) anchoring and adjustment.

Representativeness bias involves judging the probability that something will happen or an individual will have a particular characteristic on the basis of how similar the two things are. For example, subjects in a study were told that 70 engineers and 30 lawyers attended a meeting. If given no information about an individual drawn at random from the group of attendees, subjects correctly judged the probability of that individual being an engineer as .70. If told, however, that the individual drawn at random was married and well liked by colleagues, subjects judged the probability that the person was an engineer as just .50. To see the relevance for education, imagine a student transferring from another district into a middle school that offers three levels of mathematics classes. If school staff associate irrelevant personal features with mathematics difficulties, the representativeness bias could influence the student's placement.

A related fallacy discussed by Tversky and Kahneman is the failure to consider sample size when judging the likelihood of an event. When a coin is tossed, the larger the sample (the more tosses), the more likely it is that the proportion of heads to tails will be 50-50. With a small number of tosses, the more likely it is that the proportion will deviate from 50-50. In education, this fallacy is seen when achievement data for a small group rise or fall to an unusual extent on the annual testing. It is easy to assume that such an event is attributable to good or poor work on the part of school staff when, in fact, it may be the result of random variation.

Availability bias is the tendency to judge probabilities based on how easy it is to bring an example of the event to mind rather than on knowledge of base rate probabilities. Tversky and Kahneman (1982) found that this bias can lead people to make impossible probability estimates, judging the likelihood that a person chosen at random is a “shy librarian,” for example, as more likely than the chances that a person chosen at random is a librarian. In educational settings, school staff may make incorrect assumptions about the capabilities and instructional needs of particular groups of students on the basis of stereotypes or prior personal experiences with students of the same background. Subsequent studies have found that people do much better if a situation is framed in terms of frequencies rather than probabilities (Hertwig and Geigerenzer 1999). This finding would suggest that teachers will do better when reasoning with frequencies (for example, of students attaining proficiency) rather than with proportions or probabilities.

Anchoring and adjustment are a heuristic for making judgments based on initial calculations without following through on the calculations. As the complexity of necessary quantitative computation rises, people tend to anchor their final estimate on the initial solution without going through the mental labor of computation to arrive at a more accurate estimate. This mental shortcut can lead to quite inaccurate estimations in situations such as computing the effects of compound interest. In a school setting, it could lead to underestimating the cumulative effect of changes in practice that produce small effects each time they are exercised.

Extensive research on these and other decision-making biases suggests that they are partially but not completely ameliorated by decision support systems (Westbrook, Gosling, and Coiera 2005). Confidence in the correctness of a decision is not necessarily correlated with decision quality. Studies have found that decision makers have a tendency to ignore data that disagree with their preliminary decisions or biases (Batanero, Estepa, Godino, and Green 1996; Elstein, Shulman, and Sprafka 1978; Ranney et al. 2008). For example, when physicians use a data system, those who already have a preliminary diagnosis tend to pay attention to information that supports it and to disregard other information. Physicians who do not have a preliminary diagnosis in mind when they go to a decision support system, on the other hand, have difficulty knowing when they have found the answer (Westbrook, Gosling, and Coiera 2004).

Research on decision making in education suggests that educators are subject to the same biases that have been studied in basic psychology research and in other settings. Studies of district administrators making decisions have documented their difficulty in knowing what kinds of data to look for (Kennedy 1982). Education administrators have also been found to have difficulty matching patterns of data to interpretations of data that are logically coherent (Khanna, Trousdale, Penuel, and Kell 1999; Penuel, Kell, Frost, and Khanna 1998). Like medical practitioners, educators have been found to have a tendency to pay more attention to data and evidence that conform to what they expect to find (Birkeland, Murphy-Graham, and Weiss 2005; Spillane 2000; West and Rhoton 1994). In part, this is a natural response to information overload and a lack of time (Coburn, Honig, and Stein, forthcoming; Honig 2003). District administrators faced with data also reduce the cognitive load the data impose by oversimplifying them (Honig 2003; Spillane 2000). A difference between decision making within school districts and within medical practice is that decision making at the district level is socially negotiated (Coburn, Honig, and Stein, forthcoming). As noted in the research literature, the participation of multiple people including outside experts has the positive effect of mitigating some of the typical decision-making biases (Coburn, Toure, and Yamashita, forthcoming).

Much less research is available on the cognitive aspects of decision processes within schools, but there is no basis to assume that school staff are less susceptible to decision-making biases than district staff or people in general. Teachers' exercise of data literacy skills will be affected by the decision-making context and the cognitive processes and biases documented in the psychology and sociological literatures.

The next two chapters of Part I of the report examine teachers' strengths and weaknesses with respect to data concepts and skills (e.g., probability, generalizability, data computation and reduction) that can be brought to bear to reduce the biases and fallacies that often characterize human decision making. The areas in which teachers exhibit misconceptions, uncertainty, and biases in their decision processes are those in which they need more help in developing data literacy skills. To investigate these issues, the research team developed data scenarios along with a standard set of prompts that could be used to elicit teachers' thinking about student data and reveal the skills and concepts that school staff can bring to data-driven decision making. The results of administering these scenarios during school site visits, discussed in Chapters 2 and 3, shed light on how misconceptions, biases, and heuristics influence teachers as they try to make decisions based on data.

Development of the Data Scenarios

To develop the data scenarios, the research team assembled a group of internal and external experts in assessment and data-driven decision making. The group comprised two assessment experts, an expert on the use and functionalities of student data systems, a leading researcher in the area of mathematics education, and two researchers who had performed doctoral or postdoctoral research on the use of student data systems to inform educational decision making.[6]

Working with this group, the study's principal investigator identified major processes involved in using student data to inform school-level decisions: data location, data comprehension, data interpretation, data use, and question posing. For each of these components of data-driven decision making, expert group members identified specific skills and concepts that teachers should have in order to execute this aspect of data use successfully (Exhibit 1).

Exhibit 1. Data Concepts and Skills

Component	Target skills and concepts
Data location *(Finding the right data to use)*	Finds relevant data in a complex table or graph Manipulates data from a complex table or graph to support reasoning
Data comprehension *(Figuring out what the data say)*	Moves fluently between different representations of data Distinguishes between a histogram and a bar chart Interprets a contingency table Distinguishes between cross-sectional and longitudinal data
Data interpretation *(Making meaning from the data)*	Considers score distributions (not just mean or proportion above cut score) Appreciates impact of extreme scores on the mean Understands relationship between sample size and generalizability Understands concept of measurement error and variability
Data use *(Applying the data to planning instruction)*	Uses subscale and item data Understands concept of differentiating instruction based on data
Question posing *(Figuring out questions that will generate useful data)*	Aligns question with purpose and data Forms queries that lead to actionable data Appreciates value of multiple measures

After identifying these key skills and concepts for using data, the group brainstormed examples of situations or questions that would call on each of the concepts and skills. The group also reviewed screenshots from actual data systems and questions that had been used in prior research or teacher education as possible models for assessment items.

[6] These were Technical Work Group members Jeff Wayman and Ellen Mandinach, Jere Confrey, and SRI staff members Eva Chen, Geneva Haertel, and Viki Young.

The research team then used the list of priority skills and concepts the expert group developed and the example situations calling on those skills and concepts as a starting point for generating data scenarios. Each scenario described a hypothetical situation and asked the teacher to assume a certain role in that situation. Each scenario included one or more data sets and questions about what the data showed or what should be done with the data. The interview questions were designed primarily to elicit teachers' thoughts about data. Some of the questions about the data were factual, however, so that responses could be scored as right or wrong, providing an informal assessment of data literacy. An assessment expert and a mathematics education expert from the group reviewed the draft scenarios for plausibility, accuracy, and alignment with the identified skills and concepts.

The data scenarios were pilot-tested first with former teachers on the research team's staff and then with practicing teachers, with revisions made after each administration. The first wide-scale administration of the scenarios occurred during site visits at 27 schools conducted during the 2006–07 school year. On the basis of this experience, the research team revised and streamlined the scenarios for use in the second round of site visits conducted in 2007–08. Seven different scenarios were administered at site visit schools during 2007–08. To cover all the identified skill areas without extending interviews to an intolerable length, developers created two different data scenario interview forms. The amount of content on each form was balanced to achieve roughly equivalent average administration times, estimated at 30 minutes each (Exhibit 2).

Exhibit 2. Number of Items Addressing Data Literacy Components

Scenarios	Data location	Data comprehension	Data interpretation	Data use	Question posing	Total
Interview Form 1						
Scenario 1	—	—	2	1	1	4
Scenario 2	2	1	—	—	—	3
Scenario 3	2	6	5	—	—	13
Scenario 4	—	—	—	—	4	4
Total	4	7	7	1	5	24
Interview Form 2						
Scenario 5	3	2	5	—	—	10
Scenario 6	—	4	4	1	—	9
Scenario 7	—	—	—	5	2	7
Total	3	6	9	6	2	26

Exhibit reads: Scenario 1 includes two items that address data interpretation, one item that addresses data use, and one item that addresses questions posing.

NOTE: Some items address skills and concepts related to multiple data literacy components. Items are identified in terms of their primary data literacy classification.

Sample Selection

Selection of Districts and Schools

The data scenario interviews were administered as part of site visits to schools clustered within districts participating in the larger national Study of Education Data Systems and Decision Making. The case study districts were purposefully selected from a set of districts nominated by data-driven decision-making experts on the basis of their active use of student data to inform instruction. By focusing fieldwork on districts in which many teachers could be expected to be looking at student data, the research team increased the likelihood of seeing the effects of data use on practice compared with a sample of schools drawn at random. The case study districts are not a nationally representative sample. They were selected to represent best cases—sites where student data systems are available to teachers and teachers are encouraged and supported in the use of student data for decision making. These cases permitted the research team to collect information from teachers who have experience using student data and data systems. The case study site selection process began with obtaining district nominations from the project's Technical Work Group members and other leaders in educational technology, researchers, data system vendors, and staff of professional associations. These personal recommendations were supplemented with a set of districts identified through a literature search.

Ten districts obtained through this process were selected for the first round of site visits in 2006–07.[7] For the 2007–08 site visits that were the source of the data described in this report, districts that had demonstrated significant school-level data use during the 2006–07 site visits were invited to participate in a second round of data collection. In addition, more nominations of districts active in using data systems were sought from the same sources used earlier, and six additional districts were selected from this pool.

For both groups of districts, the research team worked with district administrators to identify one school that the district considered advanced in its data use practices, one school that was typical of the district in its level of data use, and one school that was emerging as a strong data user. Researchers asked the district administrators to recommend, to the extent possible, three schools serving demographically similar students at the same grade level (either elementary or middle school). The research team collected Common Core Data to characterize the sample schools and check the quality of the demographic match within districts. The case study districts included a broad range in terms of number of students served, percentage of minority students, and percentage of students who qualified for free or reduced-price lunches (poverty), urbanicity (urban, suburban, and rural), and regional location (Exhibit 3).

[7] For additional information on the district identification process, characteristics of the first set of case study districts, and findings from the initial round of fieldwork, see *Implementing Data-Informed Decision Making in Schools: Teacher Access, Supports and Use* (U.S. Department of Education 2009).

Exhibit 3. Case Study Districts and School Sample in 2007–08

District	District demographics	School type and size
District 4*	Student enrollment = 164,295 (large) Percentage minority = 50 Percentage poverty = 20 No. of schools = 238	Elementary School 1 = 732 students Elementary School 2 = 455 students Elementary School 3 = 684 students
District 12	Student enrollment = 151,421 (large) Percentage minority = 63 Percentage poverty = 40 No. of schools = 101	Middle School 1 = 2,066 students Middle School 2 = 1,082 students Middle School 3 = 1,784 students
District 11	Student enrollment = 134,002 (large) Percentage minority = 47 Percentage poverty = 28 No. of schools = 132	Middle School 1 = 1,368 students Middle School 2 = 1,105 students Middle School 3 = 490 students
District 3*	Student enrollment = 132,482 (large) Percentage minority = 74 Percentage poverty = 62 No. of schools = 219	Middle School 1 = 1,018 students Middle School 2 = 1,330 students Middle School 3 = 1,070 students
District 10	Student enrollment = 90,663 (large) Percentage minority = 83 Percentage poverty = 61 No. of schools = 89	Elementary School 1 = 464 students Elementary School 2 = 720 students Elementary School 3 = 919 students
District 1*	Student enrollment = 39,213 (large) Percentage minority = 82 Percentage poverty = 64 No. of schools = 63	Elementary School 1 = 384 students Elementary School 2 = 349 students Middle School = 585 students
District 16	Student enrollment = 27,211 (large) Percentage minority = 47 Percentage poverty = 38 No. of schools = 42	Elementary School 1 = 548 students Elementary School 2 = 502 students Elementary School 3 = 495 students
District 9*	Student enrollment = 22,174 (medium) Percentage minority = 12 Percentage poverty = 13 No. of schools = 29	Middle School 1 = 1,014 students Middle School 2 = 833 students Middle School 3 = 787 students
District 13	Student enrollment = 11,862 (medium) Percentage minority = 17 Percentage poverty = 26 No. of schools = 12	Elementary School 1 = 667 students Elementary School 2 = 537 students Elementary School 3 = 550 students
District 7*	Student enrollment = 10,780 (medium) Percentage minority = 71 Percentage poverty = 62 No. of schools = 24	Elementary School 1 = 355 students Elementary School 2 = 430 students Elementary School 3 = 339 students

Exhibit 3. Case Study Districts and School Sample in 2007–08 (concluded)

District	District demographics	School type and size
District 5*	Student enrollment = 5,599 (medium) Percentage minority = 64 Percentage poverty = 43 No. of schools = 14	Elementary School 1 = 365 students Elementary School 2 = 260 students Elementary School 3 = 399 students
District 15[†a]	Student enrollment = 1,275 (small) Percentage minority = 1 Percentage poverty = 14 No. of schools = 3	Elementary School 1 = 568 students Elementary School 2 = 316 students Middle School = 308 students

Exhibit reads: District 4 is large, with a student enrollment of 164,295 students, 50 percent of whom are minority and 20 percent of whom qualify for free or reduced-price lunches. The district contains 238 schools. Three elementary schools were included in the 2007008 site visits to this district.

*Also participated in the site visits in 2006007.

[†]Excepted from the requirement of having three schools from the same district. The Technical Work Group believed it was important to study the experiences of small districts because they serve approximately a third of public school students (Hoffman 2007). A third school from a neighboring district that the sampled district was assisting with data-driven decision making was included.

NOTES: Numbers have been used to label districts and schools for confidentiality reasons. District size categorizations (small, medium, large) are based on those for the district survey sample.

A total of 10 districts and 30 schools (19 elementary schools and 11 middle schools) were included in the 2007–08 study (Exhibit 4).

Exhibit 4. Overview of 2007–08 Data Scenario Interview Sample

Districts	10
Schools	30
Elementary	19
Middle	11
Data scenario administrations	122
Individual teacher administrations	52
Small group administrations	70

Exhibit reads: A total of 10 districts and 30 schools were included in the study. Of the 30 schools, 19 were elementary schools and 11 were middle schools. The total number of administrations of the data scenarios was 122.

Note: *Small groups* were typically composed of three teachers or of two teachers plus a school administrator, specialist, or data coach.

Selecting an Interview Sample Within Each School

During the 2006–07 site visits, individual teachers responding to earlier versions of the data scenarios had often struggled with the interview questions (U.S. Department of Education 2009). In addition, case study interviews had revealed that school staff often did much of their data use within the context of school grade-level or department teams. For both these reasons, the research team decided to collect small-group as well as individual-teacher responses to the data scenarios during the 2007–08 round of site visits. Conducting both individual and group interviews provided information about how teachers reason independently about data as well as about how they build on each other's understanding when they explore data in small groups. As part of the 2007–08 site visits, the research team administered revised data scenarios to two small groups at each participating school. Principals were asked to arrange for one small group of three teachers who would typically work together (for example, members of the same grade-level team) and one small group composed of two teachers and a school leader or data coach. In addition, at schools in districts that had not participated in the 2006007 site visits (Cohort 2 schools), the scenarios were administered individually to three teachers per school. In total, 70 small groups at 35 schools and 52 individual teachers from 18 schools completed the data scenarios.[8] The participants in the 70 small groups were 180 teachers and 35 administrators or specialists (such as data coaches). Exhibit 5 summarizes the samples for the data scenario interviews.

Exhibit 5. Data Scenario Administrations, by School Type

	Cohort 2 schools						**Cohort 1 schools (repeat site visits)**			**Grand total**
	Typical		**Emerging**		**High**		**Typical**	**Emerging**	**High**	
School level	**Indiv.**	**Grp.**	**Indiv.**	**Grp.**	**Indiv.**	**Grp.**	**Grp.**	**Grp.**	**Grp.**	
Elementary	12	8	11	8	9	6	4	10	8	76
Middle	5	4	6	4	9	6	4	4	4	46
Grand total	**17**	**12**	**17**	**12**	**18**	**12**	**8**	**14**	**12**	**122**

Exhibit reads: Of the 122 administrations of the data scenarios, 76 were with elementary school staff and 46 were with middle school staff.

NOTE: Cohort 2 schools were visited for the first time in 2007008. Cohort 1 schools were visited for a second time in 2007008.

[8] Earlier versions of the scenario interviews were used during 2006–07 case study site visits to 27 schools. Results of this round of interviews were described in a report to the U.S. Department of Education (2009).

Indiv. = Individual teacher, Grp. = Small group

The group and individual teacher responses reported in the chapters that follow represent the thinking of staff members in a particular set of schools selected from within districts that were early adopters of data use practices; they do not necessarily represent the data literacy of U.S. teachers nationally. First, the teachers were drawn from districts with a longer-than-average track record of promoting their schools' use of data for instructional improvement. Further, teachers were nominated for participation by their principals and on that basis are likely to be above average for their schools in terms of interest in data and sophistication concerning data use. On the other hand, the data scenarios presented teachers with data from hypothetical assessments in unfamiliar formats, and teachers were asked to respond to questions on demand. Both these aspects of the procedure may have tended to depress teacher performance below the level that might be expected when they are working with data from familiar assessments displayed in familiar report formats. In light of these caveats, the data in this report should be regarded as an exploration of the nature of teachers' thinking about data and how data can be used to guide instruction rather than as an estimate of the level of U.S. teachers' data literacy per se.

Data Collection Procedures

Interviewer Training

Site visitors were involved in a full-day training session that included an overview of the study's conceptual framework, the data systems each district used, and instruction on administering the data scenarios. Site visitors were shown a video of the administration of a set of data scenarios to illustrate proper administration techniques and then were given the opportunity to practice administering data scenarios to each other.

Data Scenario Administration

During the site visits, teachers participated in 45-minute interviews with two researchers. Approximately the first 15 minutes of the interview were dedicated to questions concerning the teacher's personal experience with the data system, including decisions he or she had made on the basis of student data.

Individual teachers and groups then responded to items from one of the two data scenario interview forms. Teachers and groups were randomly assigned to forms before the interview. Interview Form 1 was administered to 24 teachers individually and to 37 small groups. Interview Form 2 was administered to 28 teachers individually and to 33 small groups.

Teachers and small groups were told that the study was intended to investigate how different kinds of data displays are understood by teachers, and teachers were asked to think out loud as they looked at the various data presentations and responded to questions about them. When the data scenarios were presented, one researcher was responsible for asking teachers questions from the assigned data scenario form while the other researcher took notes.

Interviewers worked from a script with standardized questions about each data scenario. They were trained also to remind teachers to think aloud if they were silent and to ask for clarifications in cases in which a teacher's comment was ambiguous. (See Willis, 1999, for a description of think-aloud and verbal-proving interviewing approaches.)

Teachers were provided with copies of the graphs, tables, and screenshots included in each scenario. Teachers were also provided with paper, pencils, and calculators they could use as they wished (e.g., to make notations or carry out basic arithmetic calculations). All interviews were audio-recorded and transcribed to facilitate scoring and coding of teacher responses to items.

Analysis

Preparation of Transcripts for Scoring and Coding

Before scoring and coding, researchers reviewed each transcript to identify the beginning and end of the discussion of each item. Each item segment was coded with an item identification number in Atlas.ti, a qualitative data analysis program. Atlas.ti was used to produce data reports by item (i.e., all responses for a given item) to facilitate scoring all responses to the item at one time.

Scoring and Coding

Two kinds of analysis activities were conducted. First, scoring was conducted for interview questions with objective right and wrong answers to provide an indication of teachers' data literacy. Second, the entire interview transcript was coded in terms of categories related to the kind of thinking that teachers displayed with respect to the skills and concepts listed in Exhibit 1.

Scoring was done for 19 of the 24 Interview Form 1 items and 21 of the 26 Interview Form 2 items that elicited answers that could be judged as right or wrong. (For example, "What was Lake Forest School's average Total Reading Score in 2003–04?") Two raters scored each item using a detailed set of item-specific scoring criteria. (Seven of the items had multiple parts such that scores could be 0, .5, or 1; all other items were scored either 0 or 1.) Exact agreement between independent coders was 80 percent or higher for all Form 1 items. Exact agreement was 80 percent or higher for 14 of the 21 scored Form 2 items. Raters resolved all scoring discrepancies through a consensus process.

Coding categories were based on the five major data literacy skill categories and were revised and expanded during the coding of the 122 interview transcripts (Exhibit 5). Four groups of researchers (two researchers per group) received training on Atlas.ti and on coding procedures. Each group started its work by double-coding 10 transcripts and computing coder agreement for each code. Differences in assigned codes were reconciled through group discussion, and further training was provided on items with low coder agreement results. Each coder pair initiated single coding only after agreement between coders reached 80 percent on all the codes.

Contents of the Report

The next part of this report, Chapter 2, presents findings from the coded transcripts on the nature of teachers' thinking about data. Chapter 3 includes a presentation of quantitative findings concerning staff data literacy in the 35 sample schools and a comparison of the level of performance of individual teachers versus small groups. Part II of the report presents implications of these findings for teacher training and professional development and provides copies of the data scenario interviews along with guidance on issues to discuss when using them as part of professional development.

Exhibit 6. Codes Used for Transcript Scoring

Code	Definition
Action orientation	Seeks data that will suggest things teachers or school can do to control or enhance student achievement.
Alignment	Congruence between purpose question and data in a data query.
Alternative explanations	Expresses idea that other changes during this time period could be influencing performance.
Cross-sectional	Explicit mention that data represent different groups of students not the same group moving through grades.
Denominator	Attends to proportions and not just numbers proficient.
Diagnostic assessment perspective	Uses assessments that pinpoint specific areas of strength and weakness.
Differentiated instruction strategies	Describes use of assessment results to group students for differentiated instruction.
Distribution sensitivity	Examines range of the score distribution (e.g., lower, middle, and upper clusters) and not just mean score for a group.
Generalizability	Indicates sensitivity to issue of small group size, precluding generalization.
Instructional strategies	Describes strategies for improving student learning.
Item analysis	Expresses desire to get breakdown of test performance by individual test items or content standards.
Lack of knowledge (misconceptions)	Evidence that the teacher lacks basic knowledge in reading graphs, doing simple arithmetic, or interpreting simple statistics.
Logic	Refers to appropriate cell(s) in table or graph when justifying answer; answer is consistent with data or calculation based on data; must include clear conclusion and supporting evidence.
Manipulation of data	Performs mathematical operations to answer question.
Measurement error	Expresses idea that scores are based on a limited sample of observations or are only taken at one point in time.
Multiple measures	Indicates the value of using more than one outcome measure for each student before drawing conclusions.
Outlier	Comments on the effect of extreme score(s) on the average.
Perspective	Expresses idea that small differences in scale scores are not necessarily of any practical significance.
Subgroup analysis	Makes comparisons within subgroups of different ethnicities, not just comparisons of the total population.
Test validity	Understands the quality of tests. For example, standardized tests have been validated with a large number of students. A teacher-made classroom test might have limited content coverage, which can affect its validity.
Formative vs. summative assessment	Knowledge of the strengths and weaknesses of different assessments and their alignment. For example, classroom assessment can be more relevant to instructional planning.
Test fairness	Knowledge that a student’s test score can be affected by many factors including test format, the amount of test prep, availability of test accommodation, or students’ opportunity to learn.

2. The Nature of Teachers' Thinking About Data

Analyses of transcripts of case study teachers' and small groups' responses to the data scenarios provide preliminary insights into the way teachers reason with data and into the nature of misconceptions in the data literacy concepts and skills set forth in Chapter 1. Findings are described below for each of the data literacy components in Exhibit 1.

See Part II for the complete data scenario interviews, along with guidance on issues to discuss when using them as part of professional development.

Data Location

Data location skills refer to the ability to find relevant cells in a complex table or figure. Student data are typically displayed in tables, graphs, or printouts, which can be quite complex. In complex data representations, finding the desired data element is not a trivial matter. Specific skills examined within the component of data location were as follows:

- Finding relevant data in a complex table or graph
- Manipulating data from a complex table or graph to support reasoning

Finding Data in a Table or Graph

The great majority of teachers interviewed could locate specific data in a complex table or graph on request (the average percentage correct for these items ranged from 84 percent to 98 percent). For example, teachers were shown the data table in Exhibit 7 and asked to find the mean scale score of Asian or Pacific Islander fourth-grade girls who took the test. Most case study teachers (87 percent) located the relevant cell in the data table and provided the right answer (472). Case study teachers also had no difficulty in finding other types of information provided in a table or graph. For example, in responding to another question about the same table, 95 percent of case study teachers could find the number of Asian or Pacific Islander fourth-graders who took the test. Those errors that did occur in response to data location questions were generally the result of failing to correctly apply one of the requested qualifiers (e.g., student grade level or gender), with the result that the teacher read data from the wrong cell, as illustrated by the following exchange in a small group:

Exhibit 7. Grades 3–5 Students' Mathematics Scores, by Gender and Ethnicity (Hypothetical Data)

Grade	Gender	Ethnicity	Number of Students Tested	Percent of Tested Students	Mean Scale Score	Number Students at Each Proficiency Level: Below Basic	Basic	Proficient	Advanced
3	Female	African American	1	1%	589	0	0	0	1
		Asian/Pac Islander	18	26%	444	5	4	6	3
		Latino	17	24%	428	6	5	5	1
		White	34	49%	449	4	12	12	6
		Total Female	70	100%	445	15	21	23	11
	Male	African American	2	3%	452	0	1	0	1
		Asian/Pac Islander	18	23%	450	3	6	6	3
		Latino	31	40%	430	8	7	14	2
		White	27	35%	448	6	11	7	3
		Total Male	78	100%	440	17	25	27	9
4	Female	African American	2	3%	462	1	0	1	0
		Asian/Pac Islander	20	26%	472	2	7	8	3
		Latino	18	24%	441	3	8	5	2
		White	36	47%	436	8	12	12	4
		Total Female	76	100%	447	14	27	26	9
	Male	African American	0	0%	NA	0	0	0	0
		Asian/Pac Islander	16	23%	442	2	8	5	1
		Latino	24	35%	438	3	13	5	3
		White	29	42%	456	5	12	10	2
		Total Male	69	100%	446	10	33	20	6
5	Female	African American	1	1%	317	1	0	0	0
		Asian/Pac Islander	35	32%	470	6	6	8	6
		Latino	22	29%	452	4	7	8	3
		White	22	37%	470	5	8	10	5
		Total Female	80	100%	463	14	21	26	14
	Male	African American	3	4%	560	0	0	1	2
		Asian/Pac Islander	18	26%	458	4	5	5	4
		Latino	16	24%	449	2	5	6	3
		White	31	46%	464	4	12	13	2
		Total Male	68	100%	462	10	22	25	11

INTERVIEWER: What was the mean (or average) scale score for the Asian/Pacific Islander fourth-grade girls who took the test?

TEACHER 1: Well, the mean score will be 442.

TEACHER 2: No. She said female. [Crosstalk]

TEACHER 1: Oh, fourth-grade girls.

Comparing and Manipulating Numbers in a Table or Graph

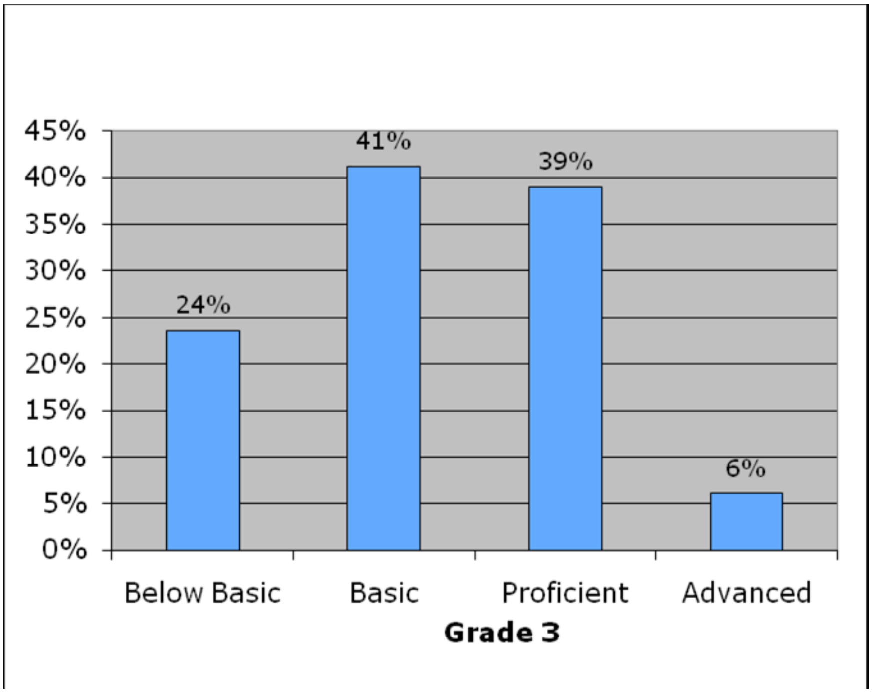
To answer some of the interview questions about data tables and graphs, teachers had to not only locate relevant information in a table or graph, but also manipulate it in some way. For example, teachers might need to compute the proportion of students with test scores below the cutoff for proficiency. Although the required numerical manipulations were simple (for example, finding a proportion), these interview questions tended to be somewhat more difficult than those that required simply finding the relevant data entry. Some case study teachers made simple mathematical errors. Other case study teachers made errors because they did not perform a needed operation on the data in the display.

Exhibit 8 shows a graph from one of the scenarios. Teachers were asked, "Based on this chart, what percentage of the school's third-graders were less than proficient in reading?"

The majority of case study participants demonstrated the ability to locate appropriate bars and perform simple calculations to obtain the correct answer. Consider the following, for example:

TEACHER: I look at the basic and I see that probably about 41 percent were basic. And if I look at the below basic it looks like about 24 percent were below basic. So when you add them together you get 65 percent were below proficient.

Exhibit 8. Grade 3 Student Reading Proficiency (Hypothetical Data)



Whether interviewed one-on-one or in small groups, the great majority of respondents correctly read the bars showing the percentages of students in the below basic and basic groups and added these numbers together correctly. Some teachers, however, focused on just the proportion of students in the basic category and ignored those who were below basic. One teacher expressed uncertainty about whether below basic students should be considered less than proficient:

INTERVIEWER: Based on this chart, what percentage of the school's third-graders were less than proficient in reading?

TEACHER: Well, it's kind of hard to say that because you got basic up here and do you add that then to the below basic too….It isn't really quite clear that these are basic and then you have to add that in also, below basic. So you could say that less than proficient would be 42 percent because that would be—also, there's another, whatever it is, 23 or 24 percent of kids that were below basic….It was difficult. You said below proficient. Then, you know, you're looking at basic and you're looking at below basic. I mean I'm not that good at reading graphs, I guess. Because I don't know if you add those two together or not. Because 40 percent is below.

A few case study teachers (8 percent) made errors in data manipulation. One teacher found and reported only below basic students as less than proficient, neglecting the students in the basic category.

In responding to questions about the table shown in Exhibit 7, some respondents (six individual teachers and small groups) appeared to be overwhelmed when asked to find the subgroup with the lowest scale score within a grade. These respondents were observed to do a partial rather than a comprehensive search of the eight relevant cells in the table and gave incorrect answers. Consider the following, for example:

INTERVIEWER: Which student group had the lowest average or mean mathematics scale score in grade 4?

TEACHER: Lowest average mean or…lowest average, or mean, mathematics scale score in grade 4?… It looks like the male Latino.

INTERVIEWER: Okay. And what was the number?

TEACHER: 438.

The research literature points to the cognitive load associated with additional information and the shortcuts that people take to reduce that load (Huang, Eades, and Hong 2009). Other research points to the role of prior expectations and ready availability of examples in decision making (Alloy and Tabachnik 1984; Tversky and Kahneman 1982). Errors may occur if respondents fail to compare all the subgroup means before giving an answer. (In this data scenario, white girls with a mean scale score of 436 rather than Latino boys had the lowest test scores among the fourth-graders).

Conclusion

Individual teachers and small groups at site visit schools encountered little difficulty in locating data in a complex table or graph. However, when asked to locate appropriate data and then perform calculations to support comparisons, some struggled either because they did not attend to key pieces of data or because they became overwhelmed by the task requirement to perform calculations and reason about the results of the calculations.

Data Comprehension

If teachers are going to make decisions based on data, they need not only to be able to find the desired data in a complex table, graph, or system interface but also to make sense of the data display. This requires a level of comprehension deeper than that needed to read a single number or compare numbers in a table or graph. In many cases, making sense of data will require teachers to reason about multiple data points from different time periods or for different entities or student subgroups. In general, data comprehension skills enable teachers to answer the question, "What do the data say?" In addition, data comprehension requires facility with a variety of commonly used data displays, which include histograms and contingency tables. Specific concepts and skills within data comprehension are the following:

- Comparing data to a verbal statement
- Understanding a histogram as distinct from a bar graph
- Interpreting a contingency table
- Distinguishing between cross-sectional and longitudinal data

Moving Fluently Between Alternative Representations of Data

To engage in thinking about and discussing data, teachers need to be able to move back and forth between tabular and graphic data representations and verbal statements about the data. In many cases, the process will require some manipulation of data in a data presentation and comparison of the data with a performance standard. The requirements of *NCLB*-inspired district and state accountability systems have created an imperative for school staff to become fluent in this skill.

One of the interview scenarios concerned grade 8 mathematics achievement in a hypothetical school whose district requires that 50 percent or more of all students and 50 percent or more of students in every student subgroup attain proficiency in order for the school to avoid designation as low performing. Teachers were shown the data in Exhibit 9 for a hypothetical school in which most of the students were Latino but some were African-American. This table displays data on the number of students, mean math score, percentage proficient, and number proficient for the two student subgroups. (In addition, teachers were told that Latino and African-American students made up the school's entire student body.)

Exhibit 9. Achievement in Grade 8 Mathematics (Hypothetical Data)

Group	Number of students	Group mean math score	Number of students proficient	Percentage proficient
Latino	239	38.5	143	60
African American	52	36.5	25	48

During the interview, teachers were asked whether, according to these data, more than half of the school's eighth-graders were proficient in eighth-grade math. To find the correct answer, teachers needed to manipulate the data by adding together the number of students who were proficient in the two subgroups, dividing the sum by the total number of eighth-grade students, and then comparing this proportion to 50 percent. Most case study teachers (75 percent) answered the question correctly, but a quarter of teachers and small groups gave an incorrect answer. In many of these cases, teachers responded not on the basis of the data but on the basis of other background knowledge or opinion. In the case below, for example, teachers misinterpreted the "percentage proficient" entries in the table as the proficiency criterion (i.e., assuming that students were judged proficient if they got 60 percent of the items on the math test correct). Their prior experience may have led them to think of a number such as 60 percent as a proficiency cutoff and to have the opinion that such a cutoff was set too low, thus leading to the conclusion that the majority of the students in the school were not proficient in math.

TEACHER 1: We don't like the standards.

INTERVIEWER: Why don't you like the standard?

TEACHER 1: It seems low.

TEACHER 2: Yes.

TEACHER 1: We would really—

TEACHER 3: Well, just by our own practice that, you know, we would—just collectively as a building or per grade level, 60 percent proficient would be extremely worrisome.

TEACHER 2: Yes.

TEACHER 3: Feel like we really need to get it better than that.

A similar question was posed about the data in Exhibit 7. Researchers asked teachers whether a majority of fifth-graders at this school had achieved proficiency in mathematics as measured by the test. To find the percentage of students who had achieved proficiency, a teacher should have added the number of fifth-grade males and fifth-grade females who were at the proficient or advanced level and then divided this number by the total number of fifth-graders at the school. Nearly a third of the case study teachers and small groups (32 percent) answered this question incorrectly. A major source of error was failure to include students classified as advanced when computing the proportion of fifth-graders who had achieved proficiency. Other teachers made mistakes in their number calculation. Finally, some seemed confused about whether they should be looking at proportions or absolute values. Several of these issues are illustrated in the transcript of a small group's members' discussion after being asked whether or not they agreed with the statement "A majority of fifth-graders at this school have achieved proficiency in mathematics as measured by this test."

TEACHER 1: Proficiency, so fifth grade.

TEACHER 2: —proficiency, no, because these two add up to—that.

TEACHER 3: We disagree.

TEACHER 2: Yes.

INTERVIEWER: You disagree? And why is that?

TEACHER 2: Because the number of students below basic and basic is greater than the number of students at a proficient level.

INTERVIEWER: Okay.

TEACHER 3: Yes.

INTERVIEWER: Next one, the—

TEACHER 2: Oh, you have the advanced. You have to consider the advanced.

TEACHER 3: But that's still not a majority.

Understanding a Histogram

The relative size of the parts of a whole can be represented graphically in several different ways. The familiar pie chart is one commonly used representation of part-whole relationships, and the histogram is another. In both cases, the percentages for the various parts should add to 100 percent. What is confusing about histograms is that their form is similar to that of a bar chart. A reader needs to attend to chart labels and understand their meaning to realize that a figure is a histogram. One of the interview scenarios was designed to probe whether teachers recognize a histogram when one is presented. Researchers showed the teachers the histogram in Exhibit 8 asked whether they saw anything wrong with it. A closer look at the histogram would reveal that the percentage of students in below basic, basic, proficient, and advanced categories adds up to more than 100. Even after being given a hint that there might be something wrong with the chart, roughly a third (32 percent) of case study teachers and small groups failed to comment on the need for the percentages in the histogram to add up to 100. They either indicated that there was nothing wrong with the chart or described some other flaw, often one irrelevant to the question or that pertained to the figure's physical appearance.

TEACHER 1: Well, the first thing I don't know the number of students on here. Secondly, there is no individual information. I also don't know that the norms are... Yes, it [the histogram] looks all right, I mean for what it is. So I don't really know.

TEACHER 2: I don't think there is something wrong with the chart. I think there is something wrong with the—I mean, naturally when I see more below basic than, or more advanced and proficient than the other two things.

Interpreting a Contingency Table

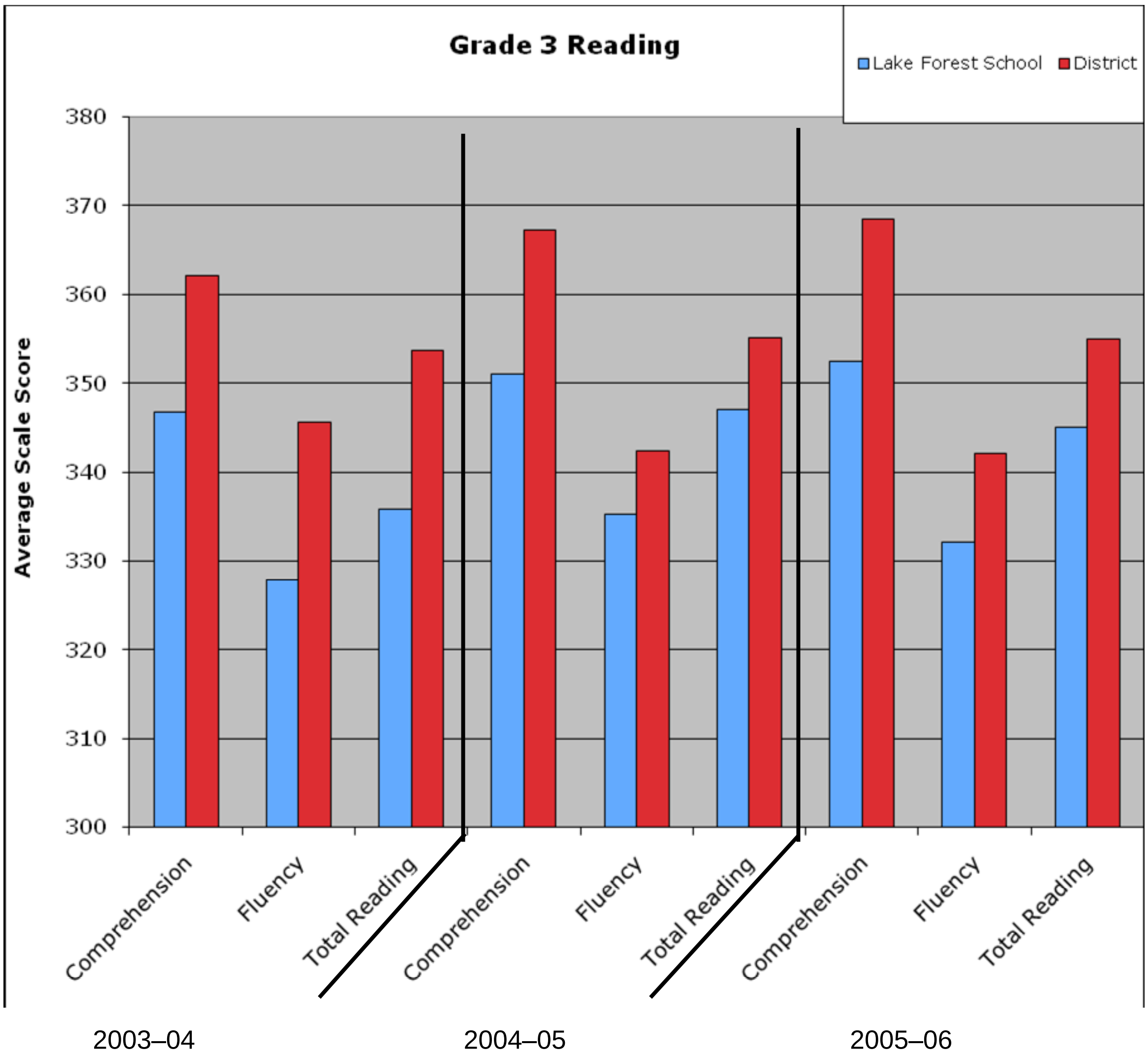
A contingency table is a way to represent the relationship between two categorical variables. The data display in Exhibit 7 incorporates a contingency table showing the relationship between student ethnicity and math proficiency status. Some of the interview questions explored the way teachers interpret data in this kind of table and their ability to understand the extent to which a relationship between the two variables is present in the data.

One of the questions teachers were asked about the data in Exhibit 7 was whether grade 5 girls were more likely than grade 5 boys to score below basic on the assessment. Answering this question correctly required finding the number of fifth-grade girls who scored below basic (14) and dividing that number by the number of fifth-grade girls who took the test (80). Then the same operation should have been conducted for the grade 5 boys (10 divided by 68) so that the two proportions (18 percent of females and 15 percent of males) could be compared. Only 51 percent of case study teachers and small groups agreed that the girls were more likely than the boys to have scored below basic on this test. Most of the teachers found the pertinent numbers in the table, but many did not take the next step of calculating percentages. Their responses could be considered an example of what Tversky and Kahneman (1982) called anchoring and adjustment. In some cases, teachers simply compared the raw numbers of students scoring below basic (14 females versus 10 males). Some teachers calculated proportions but reasoned that the 3 percent difference between the two groups (18 percent of females and 15 percent of males) was negligible and that there was really no difference between fifth-grade boys and girls. It was not possible to determine from the transcripts whether they were responding on the basis of an appreciation of measurement error or on the basis of availability bias because it was easier to bring to mind examples of girls who were struggling with math.

Distinguishing Between Cross-sectional and Longitudinal Data

Exhibit 10 shows a bar graph of hypothetical grade 3 reading achievement scores overall and separated into two components (fluency and comprehension) for a school and its district for three consecutive years. During the interview, teachers were told that Lake Forest School had started using a new reading program at the beginning of the 2004–05 school year while the rest of the district continued with the old program. By looking at the grade 3 reading scores over three years, teachers needed to agree or disagree with the statement "You can't be sure whether the program is having an effect because each year different third-graders take the reading achievement test."

Exhibit 10. Grade 3 Reading Achievement Scores Over Three Years (Hypothetical Data)



In responding to this question, 42 percent of case study teachers and small groups demonstrated understanding that the reading achievement data in the graph are cross-sectional rather than longitudinal. They reiterated the question's argument that each year a different group of third-graders took the test. They suggested that the test score means could change simply because a given year's third-graders came in with different skills. Teachers also mentioned other factors that could have affected test scores, including changes in the third-grade teaching staff or different levels of help and support from the third-grade parents or community.

INTERVIEWER: Do you agree with the statement?

TEACHER: Because they come with different abilities, and then you've got new people coming in and new teachers leaving and the way it's taught isn't the same. And the way it's received isn't the same. And the support of the parents isn't the same. And the help from the community isn't the same. So, it's just—without some other data, I can't be sure.

INTERVIEWER 2: Do you agree with the statement?

TEACHER: I would agree with that.

INTERVIEWER 2: And why would you agree with that?

TEACHER: Because they're different kids. I would like to see where you started at and where you ended at. That's my preferred…I guess, to look at where you started and where you ended because I don't care if you're still level 3 because you were in level 3 last year. That really doesn't help me out. If you're level 1 and you make it up to level 3, that means a lot more to me. So it's all based on the interpretation.

Another 42 percent of case study teachers disagreed with the statement. Even though the interview question called attention to the fact that there were different examinees each year, these teachers focused on the improvement in test scores from 2003–04 to 2004–05 without considering factors other than the introduction of the reading program.

Conclusion

The majority of case study teachers and small groups demonstrated reasonable skill in comparing data in a table or graph to a prose characterization of the data. More common were difficulties in evaluating statements about data that required calculations, recognizing a histogram as distinct from a bar graph, and recognizing the difference between cross-sectional and longitudinal data.

Data Interpretation

To make instructional decisions based on data, teachers need to go beyond comprehension per se to interpret the meaning of the data. Data interpretation requires basic data literacy skills. This is not to say that teachers need to know statistical formulas or to use statistics terminology, but they do need at least a qualitative understanding of key data concepts if they are to draw reasonable inferences from data. Data interpretation subskills addressed in the interviews were as follows:

- **Examining score distributions.** The mean score of a group of students does not provide information about individual members of the group. Teachers need to look at the information they have about individual students' skills rather than assume that a student's subgroup membership provides enough information to decide about his or her education needs.
- **Understanding the effect of outliers.** A few very high or very low scores can have a large effect on the distribution mean. Failure to look for possible outliers and take them into account can lead to data misinterpretation, especially with small data sets.
- **Appreciating limits on generalizability.** The smaller the sample (of students or of assessment items) for which data are available, the greater the risk in generalizing to other students or to other performances.
- **Understanding measurement error**. Measurement error is the difference between the obtained measurement and the "true" underlying value. Fluctuations in the state of the thing being measured and in the way it is measured can contribute to measurement error. Also, what is being measured may be probabilistic, such as the likelihood of solving an equation correctly when one is at an intermediate level of skill. Measurement error is at play every time a student is assessed. For this reason, students' scores fluctuate and multiple measures are advised.

Examining Score Distributions

Several of the scenarios provided situations designed to explore teachers' propensity to consider score distributions. In the scenario that contained Exhibit 7, for example, teachers were asked whether they thought there was a difference between third-grade boys and girls in mathematics test performance. Only about a fifth (21 percent) of the case study teachers and small groups went beyond discussing gender mean scores to examine and compare the numbers or proportions of boys and girls at each of the four proficiency levels (advanced, proficient, basic, and below basic). Similar results were found in the portion of the interview involving the school-level data in Exhibit 9. When asked if they believed both Latino and African-American student groups were doing well in math because in both cases their mean scores were above 35 (the score considered proficient), only 41 percent of case study teachers and small groups discussed the fact that even though both groups achieved a mean score above 35, not every student had a score around or above the mean. The following transcript excerpt illustrates the thinking of a group that failed to consider score distributions:

INTERVIEWER: Both of our student groups are doing pretty well in math since their mean scores are above 35.

TEACHER 2: I would agree with that.

TEACHER 1: Agree.

INTERVIEWER: And why do you agree?

TEACHER 2: Because if it's the mean, it's the average so I averaged in the low scores, too, and if it still came up as above proficient, above a 35, then I would be happy with that.

INTERVIEWER: And you—

TEACHER 1: I agree with that, agree with what she said.

INTERVIEWER: [name of Teacher 3]?

TEACHER 3: Yes?

INTERVIEWER: Grades—

TEACHER 3: Everybody has scored a—35 so they've met [proficiency].

In contrast, a different teacher demonstrated her understanding of the distinction between mean scores and score distributions in disagreeing with the statement about both student groups doing pretty well:

INTERVIEWER: Both of our student groups are doing pretty well in math since their mean scores are above 35.

TEACHER: I don't agree...Because it's still—because it's a mean score, there's still more than half the African-American kids [who] are not scoring proficient. And to be proficient you have to score at least 35. So I think the African-American kids, they may have some kids that are scoring really well, but they have more than half the kids scoring really low.

In responding to scenarios involving classroom data sets rather than school-level data, teachers were more likely to consider the distribution of scores rather than make decisions based on means. One scenario included Exhibit 11, which shows individual scores for students who had completed a unit on measurement. The interviewer asked teachers whether they would agree with a colleague who said that they should move on to the next topic in the curriculum because the class mean on the unit test was 80 percent. Nearly every case study teacher and small group (98 percent) expressed a need to examine individual student scores rather than rely on the class mean. The stark difference in teachers' attention to score distributions in Exhibit 11 compared with Exhibits 7 and 9 suggests that they are accustomed to looking at individual student scores when they have them laid out (as in Exhibit 11) but may not apply this concept to thinking about situations in which they are shown averages for an entire grade or school.

Exhibit 11. Student Scores on an End-of-Unit Examination (Hypothetical Data)

Student	Total score* (%correct)
Aaron	96
Anna	72
Beatrice	92
Bennie	68
Caitlin	92
Chantal	68
Crystal	100
Denny	88
Jaimie	68
Kayti	84
Mickey	68
Noah	96
Patricia	60
Robbie	72
Sofia	84
Stuart	68
Teresa	76
Tyler	68
Victor	100
Zoe	92
Class mean	**80.6**

*Percentage of test items the student answered correctly.

Understanding the Effect of Outliers

In responding to the data in Exhibit 11, nearly every case study teacher pointed out that about half the students in the class scored below the proficiency criterion of 80 percent. Fewer teachers discussed the potential effect of outliers on the class mean. Just 18 percent of the individual teachers and small groups responding to this scenario commented on the fact that two students had very high scores (100 percent), thus pulling up the class mean. These two teachers discussed the problem of relying on class mean scores for making instructional decisions.

The other scenarios in which outliers were included in the hypothetical data involved school-level rather than classroom-level data sets. The third-grade data in Exhibit 7, for example, included just one female African-American student who had an extremely high score (589). During the interview, teachers were asked whether there was a difference between third-grade boys and third-grade girls in mathematics test performance based on data in the table. Although male and female third-graders have similar mean scores, the extremely high score of the African-American female student pulled up the mean score for the girls. Teacher 2 in the transcript below was one of the few case study teachers (7 percent) who demonstrated an appreciation of the effect of this student's score on the girls' mean:

INTERVIEWER: Overall, based on the grade 3 data in this table, would you say there was a difference between boys and girls in mathematics test performance?

TEACHER 1: Not in the total, but an individual African-American female did better than everybody else by a large number.

TEACHER 3: But that was only one African-American female.

TEACHER 1: That's true.

TEACHER 2: But she did really, really good.

TEACHER 1: But she did better than everybody else.

TEACHER 2: Which is probably why that female is a little bit higher. The total female is a little bit higher than the total male because her score brought them up.

Appreciating Limits on Generalizability

Several of the scenarios explored teachers' understanding of the influence of sample size on the ability to generalize. One of these included Exhibit 7, which showed the mathematics scores for last year's third-graders in a school with a single African-American third-grade girl who was a very high scorer. Teachers were asked whether—assuming there were no major changes in the school's student body, teachers, or curriculum—this year's third-grade African-American girls could again be expected to outscore all the other student subgroups.

A majority of case study teachers (75 percent) commented on the fact that there was only a single African American third-grade girl in the prior year and that the score of one student is inadequate for predicting how students of the same race and gender would perform the next year.

> TEACHER: There was only one African-American girl tested in the previous year. And she had a pretty high score. I don't think that that could be representative of the whole population.

Awareness of Measurement Error

To investigate teachers' awareness of measurement error, the scenario with the end-of-unit test on measurement (see Exhibit 11) included a question about what would happen if the teacher gave the same class of students another test on measurement the following day. Every time students are tested, their scores fluctuate, even when identical test forms are used and no additional instruction is given between the tests. The great majority of case study teachers (80 percent) responded that the test results would not necessarily be the same. These teachers tended to explain their reasoning by suggesting the possibility of variations in students' state, such as feeling ill or distracted, across days. Case study teachers did not comment on the probabilistic aspect of student performance.

> INTERVIEWER: What do you think would happen if you give the same class of students another test on measurement the next day?
>
> TEACHER 1: You would still have—depends on the day, too. You're going to have some kids that—
>
> TEACHER 2: Might even have some that went down a little bit or some that went up.
>
> INTERVIEWER: Because they had a different day?
>
> TEACHER 2: Yes, different day, maybe [they are] feeling slightly different.

In responding to a similar item that asked teachers what they would expect if the *same* test were given again the next day, only 67 percent of teachers interviewed individually during the 2006–07 site visits said that the results would not necessarily be the same on a second testing. An example of a teacher response demonstrating no appreciation of measurement error was collected during the first round of site visits:

INTERVIEWER:	Okay. What do you think would happen if you gave the same class of students the same test on measurement again the next day?
TEACHER:	The next day? Nothing. They would get the same scores. There is no reteaching, there is just reassessing.

Conclusion

Most case study teachers demonstrated some understanding of measurement error. They expected student test scores to fluctuate and took into consideration situational factors that could affect student test scores on a specific day. Their concept of score fluctuations appeared to be rooted in concrete experiences with students having "off days," however, rather than in an understanding of error as intrinsic to the act of measurement.

The majority of the case study teachers appeared to understand the importance of sample size for generalizability in a variety of situations. Other data concepts, in contrast, were demonstrated by most teachers in responding to some scenarios but not others. More than one scenario contained items that measured teachers' inclination to examine score distributions, for example, and to consider the impact of an outlier on the mean score. Teachers usually examined distributions and considered the effect of extreme scores when working with classroom-level data sets but not when working with data tables or graphs that contained averages for groups of students.

Data Use for Instructional Decision Making

In order for student data and data systems to have a positive influence on student learning, teachers not only need to locate, analyze, and interpret data, but also to plan and provide differentiated instruction through techniques such as individualized learning plans, flexible grouping strategies, and alternative instructional approaches geared to different student profiles. Accordingly, the interviewers examined teacher knowledge and skills in putting data to use—that is, in identifying students' specific needs and planning instruction tailored to those needs. Three of the interview scenarios included probes of teachers' skills in making instructional decisions based on data. These items provided participants with the opportunity to demonstrate one or more of the following data use subskills:

- Understanding the value of subscale scores and item-level data
- Using student data to plan differentiated instruction based on student needs
- Synthesizing multiple data sources to inform instructional practices

Understanding the Value of Subscale Scores

The scenario associated with the grade 8 mathematics scores presented in Exhibit 9 provided an opportunity for teachers to demonstrate their interest in looking at assessment results in greater detail to understand student needs. Teachers were asked with respect to this scenario, "What actions should your school consider to avoid being labeled 'low performing' in the coming year? What other information do you need?" A slight majority of interview participants (52 percent) expressed a desire to see the breakdown of test performance by individual test items or content standards and to see individual students' performance on items or subscales in order to pinpoint students' weaknesses and adjust individual instruction.

INTERVIEWER: What actions should your school consider to avoid being labeled "low performing" in the coming year?

TEACHER 1: Okay. Well, we would need to look at the individual student data. We need to have individual data. And we need to. . .

TEACHER 2: Exactly, where, where are the missing links and, and how can we get them back up to, to par, especially in math because that's such a pyramid of knowledge.

TEACHER 1: Is there something that was a unit that was not taught well? Was there some, some particular objective that they don't understand? Building blocks, there's some building blocks missing for some of these kids.

TEACHER 2: Right. Some gaps in their mathematical progression that they can go back and fill in.

In another scenario, teachers were provided with student scores on test subscales. Given a table of hypothetical state reading test scores with vocabulary and comprehension subscales plus an overall total reading score for each student in a class, most case study teachers ignored the total reading score because they did not see much value in relying on a total score for differentiation.

Five teachers and small groups responding to this scenario (14 percent) expressed the desire for something more detailed than the subscale scores. For example, they said that text comprehension should be broken down into main ideas, sequencing, recalling, and inferring. They said that once they knew which specific strands were giving students difficulty, they could target their instruction to meet each student's needs. Similarly, a number of teachers suggested that student vocabulary test scores should be broken down further to distinguish those students who did not understand the meaning of the words from those who had difficulty spelling the words.

INTERVIEWER: Are there other kinds of information you would like to have to support your instructional planning?

TEACHER: Disaggregate some of it in the state test rather than just [the] vocabulary [subtest]. You had an idea about whether it [the students' difficulty] was the meanings of it, whether it was the spelling, whether it was something else....The same thing with [the] comprehension [subtest]. Is it [the students' problem] literal comprehension? Is it inferences?...It needs to be broken down inside.

Some teachers stressed the importance of examining both the test items themselves and students' thinking as they answered the test questions to understand why students made errors.

TEACHER: And the state achievement test, I would want to actually know how each student did. Like what I have problems with in just a vocabulary story is typically, on these achievement-type tests, you have the word and then multiple choice with meaning. So in vocabulary suggests without meaning or concept development. But it might be they just couldn't decode the word. So did they get the item wrong because it was a decoding issue or because it was word-meaning issue? The number doesn't really sort that out.

INTERVIEWER: Right. Right.

TEACHER: So that's a concern. And same with the comprehension. Sometimes students aren't as familiar with the format and how to choose a multiple-choice answer. So you don't know what the reason behind the comprehension score is. It goes back to the actual performance.

INTERVIEWER: Right.

TEACHER: You know, knowing how the child reasoned through the response to be able to find out.

Providing Differentiated Instruction Based on Data

When case study teachers were provided with individual student-level data broken down by subscale or concept, the majority demonstrated the ability to plan differentiated instruction based on data. In the scenario that included Exhibit 11, for example, teachers were provided additional data with student-level subskill breakdowns (length, weight, volume, and perimeter and area) for the end-of-unit test on measurement concepts (Exhibit 12).

Exhibit 12. Student Test Scores on Class Measurement Test (Hypothetical Data)

Student number	Length (% correct)	Weight (% correct)	Volume (% correct)	Perimeter and area (% correct)	Total score (% correct)
1	99	95	89	100	96
2	89	77	60	45	68
3	100	100	72	97	92
4	87	91	56	32	67
5	97	78	100	83	90
6	92	95	73	43	76
7	100	100	100	100	100
8	100	100	92	74	92
9	80	80	60	56	69
10	87	100	75	50	78

Teachers were asked, "Suppose that your students' performance on the various portions of the examination broke down as shown here. If you were the teacher, what would you do?" The majority of participants (59 percent) outlined strategies for providing differentiated instruction. These teachers indicated that they would provide some type of differentiated or targeted instruction through small groups or individual help in specific areas. By examining scores on subskills, teachers planned to form different groups to focus on weight, volume, and area. Some teachers also talked about pairing students up and letting students who were strong in one subskill help those who needed more practice. Other teachers planned to review the concepts in different ways, such as finding worksheets or other supporting materials that could enable students to acquire the concepts they had not yet mastered through different learning modalities. As case study teachers pointed out, most of the students performed well on length problems, except for one student. Teachers commented that this student needed one-on-one tutoring and intensive instruction focused on mastering this skill. For those students with high scores in all four subcategories, teachers planned enrichment and extension activities in measurement to challenge them so that they could integrate knowledge and engage in problem-solving activities.

> INTERVIEWER: If you were the teacher, what would you do?
>
> TEACHER: Well, I'd look at each student individually and see where they fall and then, like I said before, well, volume was a problem for certain students. Maybe I can have a small group of students that have volume problems and try to set up some learning centers. And I could say, well, this group's going to work on volume and this group, you seem to have a little bit of difficulty with weight, so let's have you go over in that learning center and try to use little learning centers where they have more hands-on to fully understand the process.

In another scenario, teachers were given the data table shown in Exhibit 13 with vocabulary, comprehension, and total reading scores on a state test as well as the results of more recent in-class assessments of sight reading and comprehension. Teachers were asked, "What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class?" More than three-quarters of the participants (77 percent) articulated different instructional content or pedagogy for groups or individual students consistent with their score profiles. These teachers and small groups described flexible grouping within the classroom to facilitate differentiated instruction based on student assessment results.

Case study teacher responses to this scenario were similar to those described for the scenario involving the test of measurement skills. Teachers typically said that they would set up one group of students for more instruction on vocabulary skills, while another group would be given lower-level texts and provided with intensive instruction on reading strategies. Some teachers talked about pairing students who were struggling with reading with good readers or bringing in reading specialists for one-on-one intervention. Many teachers stressed that their groups were very flexible and that a particular student could be in a high group for one skill but in a low group for another. As students progressed, they said that more formative assessment would be given, and students could be moved in and out of a particular group throughout the year.

Exhibit 13. Student Performance on State and Classroom Reading Tests (Hypothetical Data)

Student	2006–07 State achievement test scale score			Fall 2007 class test score	
	Total reading	vocabulary	comprehension	Sight reading	Text comprehension
Aaron	393	375	410	16	5
Anna	530	510	550	24	7
Beatrice	498	505	490	22	8
Bennie	528	515	540	26	9
Caitlin	645	660	630	28	12
Chantal	513	515	510	20	10
Crystal	573	560	585	24	10
Denny	588	566	610	20	6
Jaimie	555	550	560	25	10
Kayti	541	553	528	26	9
Mickey	410	395	425	16	5
Noah	693	678	700	30	11
Patricia	416	400	432	20	7
Robbie	563	580	545	26	8
Sofia	480	500	460	22	10
Total possible	700	700	700	30	12
Class average	530	527	532	23	8

In general, the scenario interview data suggest that teachers are more likely to think about differentiating instruction when provided with individual student-level data broken down by concept. In the scenario that included Exhibit 9 (discussed above), teachers were given only the group mean of grade 8 total mathematics scores, broken down by ethnic subgroup. When asked in connection with this scenario what actions their school should consider to avoid being labeled low performing, roughly half the respondents expressed a desire to see subscale scores, but only six teachers and small groups (10 percent) articulated the intention to provide differentiated instruction.

Synthesizing Data from Different Sources

Included in the data set presented in Exhibit 13 were several hypothetical students who had a mix of higher and lower scores on the state achievement test and the classroom assessments relative to other students in the class. This feature was designed to enable questioning that would reveal whether teachers give more credence to large-scale assessments administered for accountability purposes or to their own classroom assessments. The first question we asked the teachers was which data would be most important to them and why. Among the 65 case study teachers and small groups responding to this scenario, 45 percent indicated that both the state test and the classroom assessment data were important for making instructional decisions. Some teachers pointed out that they needed to compare students' scores on both sets of tests to see whether there were similarities or discrepancies.

INTERVIEWER: Would you consider one piece of data more important than the other or …

TEACHER: I like them both. I think I would be the more concerned about this, because I know that these kids are going to be tested again, virtually the same kind of test.

INTERVIEWER: Mm-hmm. Right.

TEACHER: And so I want to know how they're going to perform on the next year's test. I also want to know real time what they're doing right now in my opinion, because I mean, you've got [a] segment of kids that just do poorly on tests, and so I think that you want to know the real-world comprehension as well as the test.

INTERVIEWER 2: Would you consider one piece of data more important than the other?

TEACHER: I'm comparing both the comprehension scores on the state achievement test and class test. Just because I want to see that their numbers in both are similar...because if they were different, that would give me very different impressions of my students. Because there are students that score maybe really well on a state test, but then you'll give them an assessment and they won't do well.

Other teachers thought that examining both tests would reveal changes in student reading abilities over the summer and that such a recent change would be an extra piece of information to determine which students needed more intervention.

TEACHER: So I definitely look at the last year's data to see who falls below standards for last year. And then see—I just think it's interesting to see how much they lost over the summer. So I'd see who lost the most over the summer and then base my instruction on that.... So the kids who didn't lose that much, then I know they're probably reading at home and aren't going to need any interventions right away. And the kids who did lose a lot over the summer and scored low on the state assessment would need some intervention in the fall right away.

About 20 percent of the case study teachers and small groups indicated that data from classroom assessments were more important to them. The main reason for this preference was that the fall classroom assessments provided more recent data. Other teachers favored the fall assessments because of a general belief that classroom assessments, particularly one-on-one reading assessments, are more authentic, reliable, and valid. Several teachers indicated that they had limited familiarity with the items on state standardized tests.

TEACHER: If I gave it [classroom test], to me, it would matter. Did I give it individually? And therefore any kind of individual assessment, to me, when you're one on one...it is more authentic and it's more reliable than any kind of standardized test that's given to the whole class at one time... So, the validity in standardized test, that's what I have a problem with because it's not what I wanted and they're not listening to their reading and their comprehension.

TEACHER: Let's see, the part that was given by the teacher I would probably rely more on just because I wouldn't be sure, you know, I wouldn't know exactly the questions that were from the state test or the environment that they took it in. I would take it into consideration, but I would maybe compare it to see if it matched up to the assessments that I gave and if it didn't, then I might rely more heavily on mine just maybe see if I could help them with maybe it was just the format of the test or just the test itself.

Four case study teachers (13 percent) said that they would give more weight to state standardized test data in planning instruction. Among these teachers, only one explained why the state test was more useful. This teacher pointed to the broader coverage of the state test.

TEACHER: What would I use? I would use the state achievement test information. I've done the sight-reading and the text comprehension, being a reading teacher, but I think the state data provides you a broader range of sampling of the type of questions for reading comprehension, vocabulary, and overall you're going to get a bigger test range than a sight-reading score. Text comprehension, not knowing exactly what this is, but if this was one reading passage and I do a scale score and I figure out about where you would be, I know that last year this is a year's worth of education; it's 40 questions [with] some more comprehension, different types of stories in them, so I would trust this to make my adjustment with.

Several other interview questions required teachers to examine the data for individual students and make a decision about their placement in a reading group. These included a student ("Denny") who had high scores on the state assessment but performed poorly in reading comprehension on the classroom assessment. Most (80 percent) of the interviewed teachers and small groups commented on the discrepancy between Denny's high achievement test scores from the prior spring and his relatively low performance on the fall classroom assessments. Among these respondents, slightly more than half had difficulty deciding which data to rely on and what instructional strategies they should adopt for Denny. Some teachers in this group indicated they would not be able to come up with any instructional strategies until Denny took more formative reading assessment to determine his reading level.

The other teachers and small groups who commented on the inconsistencies in Denny's test results laid out a concrete plan for dealing with his future instruction. Some of these teachers speculated that Denny might have had a bad day and the classroom assessment might not reflect his reading comprehension skills. Some teachers wanted to give him another formative assessment, whereas others suggested that some one-on-one work with Denny would reveal whether he was able to read words correctly but was having difficulty understanding what he was reading.

INTERVIEWER: Well, which group would you put—say if there is a—if you put kids into small groups, which group would you put him into? Another way to ask would be, [who] are some other kids you would put into that group with Denny?

TEACHER: As a classroom teacher, I would tend to put Denny in a group that is going to have more assistance and instruction on the specific tasks that it looks like from this fall test he's lacking in. However, I would be very aware that [it] may have been a bad day. And that there are some scores from the previous year that are much higher. Now, I would also assume that I wouldn't just have this. I would have [his] report card—I would have a lot more information about the previous year. I would have a lot more information from the previous year that said he was one of the higher kids and I wouldn't be putting him in the low one. But if this is all I had, I'd be tending to go medium to medium low, but watching very carefully that I might very quickly be moving him higher

> based on the fact that there's something there that says he probably is doing much better than what this one attendance day showed.

Among the one-fifth of teachers and small groups failing to notice the inconsistency between Denny's performance on the various tests, 8 out of 11 gave a decision based on the state assessment only and three relied solely on classroom assessment data to design a lesson plan for him. For a similar interview question about how to place "Sofia" (a student with low scores on the state test but high scores on the classroom measure), nine teachers and small groups did not comment on the discrepancy; six focused only on the state assessment, which led them to put Sofia in the lowest group; and three looked only at Sofia's classroom data and decided to place her in the highest group. As a whole, responses to these questions suggest that teachers pay more attention to state achievement test data than prior case studies of teachers' attitude toward student data would suggest (Thorn 2002). It is unclear whether this finding reflects a real change in teacher attitude, the above-average stress on data in the case study districts, or simply the fact that the state assessment results appeared in the first two columns rather than the last two columns of the data display.

Analysts noted that in this scenario and the other scenarios involving data for individual students in a class (as opposed to grade-level school or district means), case study teachers tended to form a concept of individual students based not only on the data, but also on their personal experiences, a form of the availability bias discussed by Tversky and Kahneman (1982).

> INTERVIEWER: How about Denny? What group would you put him in or what approach would you try with him?
>
> TEACHER: He's very similar to a girl I have in class [in which] they can read more words. I mean, he's not at the class average as far as sight-reading, but his comprehension's so low, a lot lower than, I mean his comprehension's at 50 percent, whereas his sight-reading is not that low. So he would be in a group that would be reading like Benchmark Books in my room, but they would be easier Benchmark Books so that we could work on the comprehension. So even if he could read more words than that, it would be working on his comprehension skills.

Conclusion

The majority of interviewed teachers demonstrated their understanding of the value of examining subscale scores and conducting item analyses. When presented with student data broken down by subskills, most case study teachers described a plan for differentiated instruction based on individual student performance. However, when teachers were presented with hypothetical students with inconsistent results from different assessments, many teachers had difficulty formulating an instructional plan. There was a tendency to relate the hypothetical student to real-life students they had known and to base their decisions on experience with those earlier students rather than on data.

Question Posing

Although districts vary in their philosophy concerning direct teacher access to student data systems (U.S. Department of Education 2009), an increasing number of districts are implementing Web-based interfaces to data systems so that school staff can access and analyze data for their students. Forming a question about a set of data and expressing it as a data query is not a trivial task. The scenario-based interviews investigated three subskills in the area of question posing:

- **Aligning questions with purpose and data.** The question asked about the data set should be relevant to the goal for looking at data.
- **Forming queries that lead to actionable data.** If data use is to inform instructional decision making, it needs to shed light on options within the school's control. Actionable data are information that teachers can use to change their teaching practice. Because teachers cannot change student demographics, queries about student subgroups are examples of questions that may be unconnected to an instructional decision (unless the school is weighing the creation of a special program for one subgroup or another).
- **Appreciating the value of multiple measures.** Different tests can measure different aspects of student learning, and obtaining data from more than one source can provide teachers with a more accurate profile of students' abilities.

Aligning Questions with Purpose and Data

One scenario presented teachers with a hypothetical situation in which they had access to a computer-based student data system with the kind of interface shown in Exhibit 14. The data system contained student reading scores from state and district tests as well as semester grades in language arts. Teachers were asked to imagine that they were one of the school's fourth-grade teachers and that the school had been surprised by fourth-graders' low performance on the state reading test the prior year. Teachers were asked to describe how they might use student data from the system to inform instructional decisions that could improve student achievement.

About three-quarters of the case study teachers and small groups posed questions that aligned with their goals and the available data. Most of these teachers wanted to examine fall 2007 fourth-grade district reading assessment data first. This was the most recent reading test that these students had taken, and teachers said that analysis of these assessment results could reveal

Exhibit 14. Student Data System (Hypothetical Data)

STUDENT BACKGROUND AND ASSESSMENT INFORMATION Home

Student Information

FILTER BY VALUE: Gender Add filter

FILTER BY VALUE: FRLP Add filter

FILTER BY VALUE: Add filter

FILTER BY VALUE: Add filter

Students

Choose Student Group to Summarize:

2005-2006 Grade 3 ☐ Grade 4 ☐ Grade 5 ☐

2006-2007 Grade 3 ☐ Grade 4 ☑ Grade 5 ☐

Student Variables

Entries	Gender	Ethnicity	FRLP	Year Entered	Grade 3 Teacher	Grade 4 Teacher	2005-06 Grade 3 language arts grade semester 1	2005-06 Grade 3 language arts grade semester 2	2005-06 Grade 3 spring read achieve	2006-07 Grade 4 district fall reading
Jimmy Sampson	M	White	Yes	2004	Simpson	Kennison	462	463	436	430
Lisa Patrick	F	White	No	2003	Thompson	Kennison	481	507	448	441
Michael Scott	M	African Am	No	2003	Thompson	Ruiz	472	452	430	438
Sally Rosen	F	White	Yes	2002	Louise	Hon	430	507	436	481
Sofia Fong	F	Asian/Pac Island	No	2003	Simpson	Kennison	448	467	472	442
Tina Smith	F	African Am	Yes	2004	Thompson	Kennison	462	317	441	436
Tommy Kim	M	Asian/Pac Island	Yes	2002	Louise	Kennison	438	463	481	334
Totals	6F/5M		6 FRLP				Average 451	429	448	438

areas in which students were proficient and areas in which they needed help. Some teachers said that they would rely on this test to group students within classrooms for differentiated instruction. Many teachers also posed questions about how well the current fourth-graders (last year's grade 3 students) did on the state assessment last spring, in addition to the fall district test. They reasoned that because the goal was to raise fourth-graders' performance on upcoming state reading assessments, looking at both district and state tests could help them predict which students were likely to have difficulty. Some teachers commented that if last year's fourth-graders who scored low on the state test also scored low on last year's district tests, then the district tests could be considered practice for the state test and could be used to identify students who needed additional help before the state testing in the spring.

Although teachers' queries were coded as aligned with their purpose and available data if they involved current fourth-graders, session transcripts suggested that teachers could display this level of alignment but still have difficulty mapping between student categories in the data system and the object of their inquiry. Some teachers appeared to confuse the group of students whose performance triggered the principal's concern (the prior year's fourth-graders) with the current fourth-graders. One small group seemed to waver between this year's fourth-graders and last year's fourth-graders (this year's fifth-graders) as the target for their data investigation. Another group examined third-graders' performance the prior spring and did not realize that to see these students' fall test results they needed to look at fourth-grade, not third-grade, scores.

INTERVIEWER:	Okay. And for which group of students would you look at that data?
TEACHER:	Just the third grade because that's—that would be my student—my present students.
INTERVIEWER:	Okay. And you look at the third grade for both the state and district [tests]?
TEACHER:	Yes, like I said my, my initial [data query] would be the district but now just thinking about it I would do both [district and state test scores] just to compare how that student scored on the district and on the state.

Six case study teachers and small groups failed to identify the student group portion of their data query at all. They talked about the measures they would like to examine but not the student groups for which they would get those measures in a data report.

Forming Queries that Lead to Actionable Data

Teachers can use actionable data to adjust their teaching practice in ways that enhance student learning. For example, item-level analysis of assessment results or student subscale scores in reading provides information that teachers can use to put greater or less emphasis on certain topics or to plan individualized instruction for students. Reports of student performance by socioeconomic status or ethnicity will not necessarily provide data suitable for teachers to act on to improve student learning. The scenario associated with Exhibit 14 asked teachers to think of data that they might want to help improve fourth-graders' reading achievement.

Roughly two-fifths of case study teachers (42 percent) indicated that they would want data that analysts considered "actionable" in nature. The majority of these teachers expressed a desire for subscale scores or item-level data from the state and district tests. They did so even though the hypothetical system interface shown to them did not offer access to this kind of information. Their goal was to investigate the areas in which their students were particularly weak so that they could plan grouping and reteaching. As illustrated in the words of two different teachers below, case study teachers appear to have become quite savvy regarding the payoff of matching instructional coverage to the content covered on a state test.

TEACHER: Well, I'd want to look at what the fourth-graders did on the state test. I'd want to look at what the third-graders did on the state test and look at the areas of weakness and strength and then also look at the test itself, because it's sort of what percentage of what strand is on the test? It's not just one big thing. It's different categories on the test. Basically start to focus on what's the biggest percentage of the test and then where do these kids fit in as far as strengths or weakness on those different percentages?

TEACHER: Because we'd want to see if—'cause that's the test they scored low on, so that's where I want to start.... What particular part of that test did they score low on? That's what I'd want to know.... Whether there is one certain area—was it vocabulary, was it comprehension? You know, what areas did they score low on.... And what would I do with it?... I'd break it down. I'd love to have an item analysis... and go back even if you have to go back through old lesson plans and old assessments and see how you were teaching the skills and knowledge that they seem to kind of fall down on and how you were assessing those skills and knowledge and make some decisions about some changes that might need to happen.

The other kind of data query classified as "actionable" had to do with obtaining student assessment results disaggregated by teacher. Case studies of data use in schools have found that at schools that are more advanced in using data to improve instruction, small groups disaggregate student data by teacher in order to identify the teachers with the strongest student results so that they can learn from those teachers' practices (U.S. Department of Education 2010). About a quarter of the case study teachers and small groups responding to this data scenario indicated that they would want reports identifying fourth-grade students' former (third-grade) teachers. Teachers noted that this would enable them to obtain student profiles directly from the former teachers. They stated that they often find information learned from past teachers valuable in helping them design differentiated instruction. Furthermore, disaggregating student data by teacher might reveal specific student performance patterns associated with different teaching approaches adopted by various teachers. For example, if a teacher discovered that a group of students in her class who were taught by the same teacher last year all performed well, she could talk with the former teacher about the strategies she used.

TEACHER 1: And also look at the teachers, [how] they graded them and maybe talk to the teacher from the year before.

TEACHER 2: Right.

TEACHER 3: I think this data could be potentially valuable, too, if you see a particular problem that just might pop up. You might not anticipate a particular problem. If you had access to any of this, you might see a pattern.

It was less common for case study teachers (role-playing the part of a fourth-grade teacher) to say that they wanted to disaggregate last year's fourth-grade state assessment results by fourth-grade teacher. An exception is provided below.

TEACHER: I think you also want to look at a teacher. Because if you are disaggregating this data, you want to see if any kids with a certain teacher performed better or worse. And you know, use that person as a resource or identify, you know, maybe they just didn't cover a unit or something happened. Or maybe it was a traumatic event. I mean you would kind of want to look at that data as well.

Other teachers and small groups focused on disaggregating data by variables that were beyond their control. Thirty-five percent of case study teachers and small groups responding to this scenario said that they would disaggregate the achievement results by student background variables. These teachers formed queries concerning student performance by gender, socioeconomic status, language background, and special education status. They noted that these factors are related to student achievement but were unable to explain how they would act on those data to inform their instruction.

INTERVIEWER: So just looking at the data that you have in front of you right now, is there anything else that—any other questions that you want to ask that can be answered by this data?

TEACHER: Well, by looking at the free or reduced-priced lunch [status] you have some idea as to, you know, where your kids are coming from, which unfortunately can tend to have a pretty significant impact on test scores. You wouldn't want it to, but it does.

Understanding the Value of Multiple Measures

In the interview that included Exhibit 14, teachers were asked whether there were other data they would want to see represented in the data system. Only one of the six case study teachers responding to this scenario (17 percent) made an explicit statement about the importance of having multiple measures to better understand students' strengths and weaknesses. However, many teachers (37 percent of case study teachers and small groups responding to this scenario) articulated the value of using classroom activity, including one-on-one reading work with students, as a basis for understanding students' skills and needs. Some teachers expressed the need to examine student scores on formative reading tests. Others indicated that in addition to state and district test scores, teachers could learn a lot more about students by looking at a portfolio of their work.

TEACHER 1: Yes, more formative reading assessment, not just a state or district assessment. I want to do IRIs, DRAs, individual—all of those reading assessments.

TEACHER 2: All those pieces.

TEACHER 2: All those pieces that break up and tell you more information about them [students].

TEACHER 1: Specific strategies and then informal, guided readings. I would sit with them and what you can pick out from listening to them read.

Education data systems have been criticized for their lack of information on the educational experiences that specific students have had (Besterfield-Sacre and Halverson 2007). It is difficult for schools and teachers to evaluate the effectiveness of the different things they do if they cannot see outcome data related to programs and interventions. A few case study teachers expressed a desire to see this kind of information in the data system.

> TEACHER: I'd like to know if they were in after-school remediation, if that played a part in things. A lot of times kids can do a whole lot better just by doing that. If they went to summer school, that wouldn't be on this type of thing, but all those things are very helpful to know.

Conclusion

Among the question-posing subskills, looking for multiple measures to inform decisions appeared to be the one case study teachers most widely demonstrated. In many cases, teachers wanted data from assessment subscales or item analyses, which are forms of actionable data. Still, over a third of case study teachers and small groups responding to this scenario described planned data queries that were irrelevant to the goal of raising their current fourth-graders' reading achievement or that concerned demographic (status) variables beyond the school's influence.

3. Data Literacy Among School Staff

Chapter 2 provided descriptive data on how case study school staff thought about the data scenarios and responded to interviewer probes. This chapter reports on those portions of the data scenario interviews that could be scored as correct or incorrect. An examination of the frequency of correct performance on interview items related to the various components of data literacy provides insights into areas that need additional attention in teacher preparation and professional development programs.

See Part II for the complete data scenario interviews along with guidance on issues to discuss when using scenarios as part of professional development.

Data Literacy at the School Level

To estimate data literacy at case study schools, the research team analyzed scale scores at the school (rather than the individual teacher) level. The scores for all the teachers at a school who took the first item were averaged, and that mean score was assigned to the school; this was repeated for the second item, and so on. Through this process, the data set was structured as 35 school-level records with scores for items on both forms.

The research team estimated reliability by computing an alpha coefficient for the total score. Despite the small school sample and the intentional inclusion of distinct abilities in the assessment, the reliability for the total scale, including eight items from Form 1 and nine items from Form 2, was fairly high, alpha = .71. The items included in the total score represent all five of the hypothesized components of data-driven decision making. The total score mean was .71. Mean scores on component subscales suggest that question posing and data interpretation are the most difficult data literacy skills for teachers (Exhibit 15). However, some of the subscales representing individual components of data literacy (e.g., question posing) were not highly reliable, so subscale results should be used only with caution.

Exhibit 15. Frequency Distribution and Mean for Scored Items

	Number of items*	Mean number of respondents per item	Mean score	Standard error	Alpha of scale
Total score	17	58	.71	.02	.71
Data location	4	58	.92	.02	.54
Data comprehension	5	59	.64	.03	.62
Data interpretation	4	59	.47	.04	.61
Data use	3	53	.64	.04	.40
Question posing	3	48	.34	.04	.35

Exhibit reads: Teachers' average score across the 17 data scenario items included in the total score was .71. The alpha for the total score scale was .71.

*The number of items associated with the total score and components is less than the total number of items administered. Items were removed from scales to improve the reliability of the scales.

Data-informed Decision Making by Groups Versus Individuals

Teacher survey data indicate that data-driven decision-making activities are as likely to be conducted in groups as they are individually (U.S. Department of Education 2009). The first round of site visits to case study schools indicated that teachers in these schools often worked together in grade-level or subject-area teams to examine student data. The performance of individual teachers responding to the pilot data scenarios as part of the first round of site visits was not high (see U.S. Department of Education 2009), but the research team reasoned that teachers might demonstrate stronger data literacy skills working in small groups. To explore this issue during the second round of site visits conducted in 2007–08, researchers administered the data scenarios to both small groups and individual teachers within the 18 schools visited for the first time that year (Cohort 2 schools).

Exhibit 16. Cohort 2 School Means for Individual Teachers and Small Groups

	Individual teachers		Small groups	
	Mean score	Mean number responses per item	Mean score	Mean number responses per item
Total score	.64	25	.72	17
Data location	.85	24	.97	17
Data comprehension	.52	26	.69	18
Data interpretation	.60	26	.59	17
Data use	.62	28	.58	12
Question posing	.16	19	.43	13

Exhibit reads: Individual teachers in Cohort 2 schools had an average score of .64 across the 17 data scenario items compared with a score of .72 for small groups in the same schools.

The data literacy total scores are based on 17 items, and scores for groups were higher than scores for individual teachers on 13 items (Exhibit 16). The difference between group and individual scores was significant for 5 items. This performance pattern suggests that small groups of teachers may be able to extract more useful information from student data sets than teachers working in isolation.

The research team conducted qualitative analyses of transcripts on the five items for which significant performance differences existed between groups and individual teachers. Analysts first examined the degree to which groups and individuals discussed the same subskills related to each component of data-informed decision making. They then looked at differences in the problem-solving process of groups and individuals in terms of how they framed the problem and solution and their affect during problem solving.

Skills Demonstrated by Small Groups and Individual Teachers

For each item, coders looked for the presence of specific concepts or skills that demonstrate the five components of data-informed decision making.

Two of the five scored items that groups dealt with more successfully than individual teachers involved the scenario showing three years of reading achievement data for a school and its district. One of the items was whether respondents agreed with the following statement: “Lake Forest School’s progress in narrowing the grade 3 reading achievement gap compared with the rest of the district has been in reading fluency rather than reading comprehension.” Groups were more likely than individual teachers to manipulate the data to compare them with the verbal statement. The groups were more likely to compute the difference between the reading fluency scores in the first and third years and between the reading comprehension scores in the same years before deciding whether they agreed or disagreed with the verbal characterization of the data (20 percent of case study groups did so compared with 13 percent of individual teachers).

The other item from this scenario with a significant difference asked teachers whether they agreed with the statement “Lake Forest School has not benefited from the new reading program since it was first implemented in 2004–2005.” The greatest difference for this item was that 19 percent of the groups talked about possible alternative explanations for the rise in the school’s scores in 2004–05, whereas only 11 percent of individual teachers did.

Teachers were asked whether they agreed with the following statement based on data from the 2006–07 school year: “This year’s third-grade African American girls will score better than other students when this test is given to this year’s third-graders.” The research team analyzed transcripts to determine whether teachers referred to appropriate cells in the table when justifying their answer and included a clear conclusion with logical evidence, whether teachers indicated a sensitivity to the issue of small group size precluding generalization, and whether teachers manipulated data to support their reasoning about a verbal statement and explicitly mentioned that data represent different groups of students each year. The major difference between group and individual transcripts on this item was that a majority of groups (55 percent) discussed the hazards of generalizing from a single student in one year to a new cohort of students the next year, but only 29 percent of individual teachers did.

One of the scenarios showed teachers an interface for a hypothetical electronic data system and asked them what achievement data they would want to look at to inform their instruction in ways likely to improve fourth-graders’ reading performance. There was a significant difference between groups and individual teachers on the follow-up questions, “Would you like to make any other queries of the data system? Are there any other questions you want to ask that can be answered by the data in the system?” Transcripts were analyzed to evaluate whether teachers identified questions that were congruent with the data, indicated the value of using more than one outcome measure for each student before drawing conclusions, expressed the desire to get a breakdown of test performance by individual test items or content standards, sought data that would suggest things teachers or school could do to enhance student achievement, and mentioned background information about students that should be considered. Case study small groups were somewhat more likely than individual teachers to express a desire to be able to do an item analysis on the state test (28 percent versus 20 percent). Groups also were more likely than individual teachers to talk about background variables that might affect achievement results (45 percent compared with 20 percent). These included both demographic variables and prior educational experiences. Consider the following, for example:

TEACHER:	Well sometimes, see, the year that they entered the district, because different, you know, districts have different programs and things like that.
INTERVIEWER:	I see.
TEACHER:	I know, like, when we get a child from another school district, that kind of—we always want to know what they've learned over there and—
INTERVIEWER:	Okay. So if they're coming from a different district, how long they've been in the district, and things like that, I see.
TEACHER:	Right.
INTERVIEWER:	So their background knowledge might be different.
TEACHER:	Right.

In a few cases, groups of teachers considered how they might use background information to identify additional learning support for students, as illustrated in the example below.

TEACHER 3:	IEP.
INTERVIEWER:	Yes.
TEACHER 2:	Yes. So you could see how much help they had, or if they had any help, or they needed help but they couldn't get it.
TEACHER 3:	Right.
TEACHER 1:	I would probably want a little bit more actually. I would probably be more interested than [Teacher 2] was in seeing if there were any patterns in some of those background variables, because from a wider building perspective, there might be some other things that we can do differently to prevent these situations from happening again. We might be not doing a very good job of meeting their needs in various, maybe it's an ethnic thing— [we] want to be fair. Or maybe it's students that have only been in the district a couple of years. Maybe they're the ones lagging behind their peers and maybe we can do something to address that in the future.

Problem-solving Approach by Small Groups and Individual Teachers

Differences in problem solving by case study small groups and individual teachers were notable in terms of problem clarification, error identification, and affect.

Problem Clarification. Teachers working in groups have the opportunity to clarify and discuss how to interpret the question and frame the problem. An individual teacher's misreading of the data (e.g., confusing trends in absolute performance with changes in performance relative to a larger population) can be corrected by a colleague. Clarification allows groups to use evidence to make more appropriate decisions.

INTERVIEWER: Lake Forest School's progress in narrowing the grade 3 reading achievement . . . gap compared with the rest of the district has been in reading fluency rather than reading comprehension. Do you agree or disagree?

TEACHER 1: I agree with that—

TEACHER 3: Fluency, fluency, fluency, and then the district—

TEACHER 1: Yeah, but it's still overall.

TEACHER 3: Fluency overall, I don't know. District fluency up.

TEACHER 1: The district—

TEACHER 3: It looks like—

TEACHER 1: The district fluency went down every year.

TEACHER 2: So reading fluency and reading comprehension. Fluency is—in the middle.

TEACHER 2: It looks like, to me—

TEACHER 3: But the district didn't make as much progress as—I mean the school didn't—

TEACHER 2: But they are saying—they are not talking—they are talking about narrowing.

TEACHER 3: Narrow the gap.

TEACHER 2: So that you want the lines to look getting smaller.

TEACHER 1: Right. It was a big gap in '03–'04 in fluency.

TEACHER 2: It was a big gap, yes. It went up.

TEACHER 1: And then ‘04–‘05, there is a—quite a significant change because the district’s fluency went down so low.

Error Identification. Teachers who work in groups have the opportunity to catch mistakes made by their colleagues, so groups may be more likely to reach accurate conclusions.

INTERVIEWER: Okay, what about this question? Based on this chart, what percentage of this school’s third-graders were less than proficient in reading?

TEACHER 1: 30.

TEACHER 2: Less than proficient?

INTERVIEWER: Less than proficient.

TEACHER 1: Oh, I'm sorry. Sorry, no.

[Crosstalk Discussion]

TEACHER 1: 71.

TEACHER 2: 65.

TEACHER 1: 65, I'm sorry. [Laughter]

INTERVIEWER: And what was your process for getting to 65?

TEACHER 2: I just estimated the basic.

INTERVIEWER: Yes.

TEACHER 2: And I estimated the below basic, and I added them together.

TEACHER 1: For some reason I was adding the advanced, and that’s where my 71 came from. I was totally out to lunch there.

Affect. The experience of making data-driven decisions may be more enjoyable for teachers if they work in small groups. Coders working with the data scenario transcripts noticed a difference in the emotional tone of the individual versus the group discussions of data. To get an objective behavioral indicator of affect, analysts coded the number of times laughter was noted in the transcripts for the five items on which individual teachers and small groups differed in response quality. Laughter was noted in 21 percent of transcripts of the small groups compared with just 9 percent of transcripts of individual teachers.

Conclusion

In summary, working in small groups appears to promote several aspects of teachers' engagement with student data. Groups are not only more likely to arrive at sound data interpretations but also appear to use a wider array of skills to inform decisions about how to interpret and use data when compared with individual teachers. Working in groups may afford teachers the advantages of clarifying and framing problems and correcting data interpretation errors with help from colleagues. Finally, the researchers' observation that case study small groups seemed to enjoy the process of analyzing and interpreting data more than teachers working alone suggests that opportunities to use data with colleagues may be easier to scale and sustain than policies that rely entirely on individual data use.

4. Discussion and Conclusion

The use of student data to improve instruction is a central tenet of current education policy (*American Recovery and Reinvestment Act* 2009). Various accountability mandates including those in *ESEA* stress the importance of data-informed decision making. Current efforts to improve school performance are calling on teachers to base their instructional decisions on data. More and more, teachers are expected to assess students frequently and to use a wide variety of assessment data in making decisions about their teaching (Hamilton et al. 2009; Schmoker 1996; U.S. Department of Education 2004).

Teachers' Data Skills

Notwithstanding its exploratory nature, the study described here demonstrates that student data do not speak for themselves. Even within districts such as those in these case studies, with a reputation for supporting data-driven decision making, some teachers struggled to make sense of the data representations in the assessment interviews. Especially when the question called for framing queries for data systems or making sense of differences or trends, a sizable proportion of case study teachers made invalid inferences. The most difficult data literacy concepts and skills appeared to be reasoning about data when multiple calculations were required, interpreting a contingency table, distinguishing a histogram from a bar graph, and recognizing differences between longitudinal and cross-sectional data. For example, most case study teachers could compare tabular or graphic representations with verbal descriptions fairly well, but some compared raw numbers when responding to statements about proportions. It is unlikely that teachers in districts as a whole, most of which have put less emphasis on teachers' use of data, would have less difficulty than the teachers whose responses are described here.

Teachers also displayed many of the decision-making heuristics and resulting biases studied in experimental situations by Tversky and Kahneman (1982) and in naturalistic studies of district office decision making by social organizational researchers (Birkeland, Murphy-Graham, and Weiss 2005; Coburn, Honig, and Stein in press; Spillane 2000). Particularly in dealing with scenarios involving grade- or school-level data, some case study teachers appeared to lose track of what they were trying to figure out. Other teachers started making calculations but then ceased using a numerical approach, instead relying on their general impression to answer the question if the calculation became at all complicated (a tendency Tversky and Kahneman describe as the anchoring and adjustment heuristic).

When given an open-ended invitation to explore data for the purpose of improving achievement, teachers had difficulty defining clear questions and did not ask questions that could eliminate rival hypotheses. For example, few case study teachers wanted to look at a school's grade-level achievement data by teacher, which could have provided some insight into whether some teachers' practices were less effective than others'. It was also rare for interviewed teachers to ask about comparing successive cohorts of students on the same outcome measure (to see whether the prior year's unexpected low performance was likely to be related to a specific cohort).

In general, case study teachers were both more comfortable and more adept when dealing with familiar situations, such as the interpretation of results of a classroom assessment. Regardless of the kind of data presented (classroom versus school- or district-level data) and the different settings (working individually versus in a small group), most case study teachers demonstrated skill in the familiar tasks of data location and data use for instructional planning. However, in the more challenging skill areas of data comprehension, data interpretation, and data query, teachers performed better when interpreting classroom-level data or in interpreting school- or district-level data with the input from their colleagues. Case study teachers had the most difficulty with data comprehension, data interpretation, and data query when they worked individually with summative assessment data.

Influence of the Type of Data

With class-level data, case study teachers tended to examine distributions and look for outliers in a way they did not when given grade-level or school averages. At the same time, however, it was fairly common for case study teachers to use particular past experiences with individual students as a basis for forming decisions about the hypothetical students in the data set. Although there are certainly advantages of experience, the tendency to respond to a new student with a strategy that was effective with a past student can be overused, and one of the intended advantages of data-driven decision making is to ensure that objective criteria rather than intuition or demographic stereotypes are the basis for instructional decisions.

Influence of the Social Setting

A comparison of teacher responses during individual and group administration of the data scenarios suggests that teachers are more likely to reach valid conclusions and exhibit a broader range of data literacy concepts and skills when working in small groups. Interviews of teachers at case study sites (U.S. Department of Education 2010) provide further support for the inference that group interactions support teachers' use of data and application of data to improving their instruction. Especially during this period when data use for improving instruction is still a new activity for most teachers, there may be advantages to providing the social and intellectual support that can come from working in groups.

Next Steps

More than 230 teachers participated in this exploratory research, either in individual interviews or in small groups. This report describes detailed examples of the way they interacted with student data and attempted to make sense of the data. This exploratory work provides an initial look at how teachers reason with student data.

Need for Further Research and Development

Given the importance that federal education policy places on teachers' data literacy and ability to draw instructional implications from data, additional research and development are needed to move this work from the exploratory to the operational phase.

An expanded item bank with additional hypothetical data sets and queries is needed to fill out the data literacy component subscales and increase their reliability. With an expanded set of data literacy items in hand, researchers could validate the items against measures based on observations of teachers' use of real data from students in their own classrooms and schools. Combining assessment and ethnographic research, such a study could answer the question of whether teachers who perform better with the hypothetical data scenarios also make more use of student data and make better decisions based on data in their daily practice. The same study could also address the question of whether data use skills and concepts important in everyday practice are missing from the five data literacy components and associated skills that provided the framework for developing the data scenarios used in this exploratory research.

Having a validated teacher data literacy assessment would then enable additional research and evaluation. Such a measure would be extremely useful, for example, in evaluating teacher preparation and professional development activities intended to prepare teachers for data-driven decision making. In addition, the National Center for Education Statistics might want to consider administering validated assessment items to a nationally representative sample of teachers to provide a national snapshot of teacher competencies in this arena. Administration of such a data literacy assessment at multiple time points could provide an indication of the extent to which federal, state, and local policies are fostering these skills in the U.S. teacher workforce.

Finally, as noted earlier, the rationale for this exploratory study came from national survey data showing that both district administrators and teachers themselves express reservations about teachers' ability to make sense of student data reports provided by electronic systems. The national data were collected from districts during the 2007–08 school year and from teachers during the 2006–07 school year. As policymakers continue to emphasize data use to support education reforms, it will be important to track progress made nationally in teachers' data literacy and ability to use data.

Implications for Policy and Practice

Fulfilling the national policy mandate for data-driven decision making will require teachers to acquire a deeper understanding of basic assessment and statistical concepts and to become fluent in reading various data representations (tables, charts, dashboards, database interfaces). Teacher preparation programs and district professional development offerings both are needed in this area. Districts are concerned about teachers' lack of preparation in how to use data for instructional decision making. In a national district survey, 72 percent of districts cited lack of teacher preparation as a barrier to increased use of data systems (U.S. Department of Education 2008). Teachers also express a need for professional development related to the use of data to shape instruction. On a national teacher survey, 58 percent of teachers said that they could benefit from additional professional development on how to develop diagnostic assessments for their class, 55 percent on how to adjust the content and approach used in their class in light of student data, 50 percent on how to identify types of data to collect in order to monitor school progress against goals for improvement, 48 percent on the proper interpretation of test score data, and 38 percent on how to formulate questions that can be addressed by data.

Both research on best practices in professional development (Adelman et al. 2002; Porter, Garet, Desimone, Yoon, and Birman 2000; Lawless and Pellegrino 2007) and the accounts that teachers gave of their data use practices in the Study of Education Data Systems and Decision Making case studies (U.S. Department of Education 2010) suggest that the best support for acquiring and honing data interpretation skills will be sustained participation in teacher learning communities using student data. (See also Hamilton et al. 2009 for a similar conclusion.) The exploratory work reported here is consistent with this hypothesis. When working with data in groups, case study teachers were more likely to attend to relevant information and to have their tendency to arrive at answers without sufficient analysis challenged by colleagues. Teacher learning communities also are well situated to go beyond the interpretation of the data per se to consider the instructional options for dealing with areas of difficulty. Only by bringing insights derived from student data together with appropriate instructional strategies will teachers be able to achieve desired improvements in student learning.

Part II. Resources for Data Literacy Professional Development

5. Using the Data Scenarios for Teacher Professional Development

The remainder of this report presents the data scenarios used in this research, along with commentary on what to look for in teachers' responses to them. The scenarios can be used to acquaint teachers with different kinds of data representations and different data interpretation issues, without invoking the anxiety and defensiveness that sometimes arise when teachers look at data from their own classrooms. Facilitation of the discussion of the scenarios by a data coach or training is recommended, especially for scenarios that involve data literacy concepts that are difficult for many teachers. Exhibit 17 lists the seven data scenarios with a brief description of each and an explanation of the skills and concepts addressed.

These scenarios can be used as the focus for teacher professional development activities. Teachers can be organized into small groups with each group discussing a scenario and responding to the questions. After a scenario has been discussed in small groups, a trainer, coach, or school leader can facilitate a joint discussion of the scenario, covering the skills and concepts highlighted with respect to each question in the scenario interview, as shown below. Facilitators with a background in assessment and data analysis will be able to use the discussions prompted by the scenarios as a jumping off point for a deeper treatment of data literacy concepts.

These scenarios are intended as a resource for discussion. There is no implied claim that learning to respond correctly to the scenarios will by itself change teacher practice or enhance student achievement.

Exhibit 17. Data Scenario Interviews

Scenario	Overview	Skills and Concepts Addressed
1	The teacher's sixth-grade class has completed a unit on measurement and taken an end-of-unit test. Given the class average and the distribution of scores for the 20 students in the class, the teacher must decide whether to move on to the next math topic or reteach some of the measurement content to some or all students.	Considers score distributions (DI) Appreciates impact of extreme scores on the mean (DI) Uses subscale and item data (DU) Understands concept of measurement error and variability (DI) Understands concept of differentiating instruction based on data (DU) Appreciates value of multiple measures (QP)
2	A histogram shows a school's third-graders' proficiency classifications on the state reading test. Teachers are asked to identify the proportion of students achieving proficiency and to examine the display for a possible error.	Finds relevant data in a complex graph (DL) Manipulates data from a complex graph to support reasoning (DL) Distinguishes between a histogram and a bar chart (DC)
3	Teachers are shown a bar graph with three consecutive years of reading achievement data for a school and its district. They are asked to compare school trends to district trends and to consider whether a new reading program implemented at the school in the second year is proving effective.	Finds relevant data in a complex graph (DL) Manipulates data from a complex graph to support reasoning (DL) Moves fluently between different representations of data (DC) Understands concept of measurement error and variability (DI) Uses subscale and item data (DU) Appreciates value of multiple measures (QP) Distinguishes between cross-sectional and longitudinal data (DC)
4	Teachers are asked to suppose that they have access to an electronic data system containing spring state reading test scores, scores on district tests administered in the fall, English language arts grades, student demographics, and teacher names. They are asked what data they would like to see if formulating a plan to improve fourth- graders reading scores.	Aligns question with purpose and data (QP) Forms queries that lead to actionable data (QP) Appreciates value of multiple measures (QP) Uses subscale and item data (DU)

See note at end of table.

Exhibit 17. Data Scenario Interviews (concluded)

Scenario	Overview	Skills and Concepts Addressed
5	Given a table showing mathematics achievement data for grades 3, 4 and 5 by student subgroup, teachers are asked to find specific information in the table and to use the table to determine whether a series of statements about the students' mathematics performance are true or false.	Finds relevant data in a complex table (DL) Manipulates data from a complex table to support reasoning (DL) Interprets a contingency table (DC) Considers score distributions (DI) Appreciates impact of extreme scores on the mean (DI) Understands concept of measurement error and variability (DI) Understands relationship between sample size and generalizability (DI) Moves fluently between different representations of data (DC)
6	Teachers are asked to consider their schools' math achievement data in terms of district requirements for the proportion of students in each subgroup meeting a set proficiency requirement.	Manipulates data from a complex table to support reasoning (DL) Moves fluently between different representations of data (DC) Finds relevant data in a complex table (DL) Considers score distributions (DI) Appreciates value of multiple measures (QP) Uses subscale and item data (DU) Understands concept of differentiating instruction based on data (DU) Understands relationship between sample size and generalizability (DI) Understands concept of measurement error and variability (DI)
7	A table shows reading achievement test scores and scores on two classroom assessments for 15 students. Teachers are asked what instructional decisions they would make based on this data and how they would place several students who have high scores on some measures and low scores on others.	Uses subscale and item data (DU) Understands concept of differentiating instruction based on data (DU) Appreciates value of multiple measures (QP) Understands concept of measurement error and variability (DI)

Note: Components of data literacy are identified in parentheses next to each skill or concept: DL = Data Location, DC = Data Comprehension, DI = Data Interpretation, DU = Data Use, QP = Question Posing.

6. Data Scenarios

Below are the actual scenarios as presented to respondents.

SCENARIO 1

SCENARIO: Now I'm going to show you some data from a hypothetical classroom.

Suppose you're teaching mathematics to a class of 20 sixth grade students and at the end of a unit on measurement, you gave a 100-point multiple-choice test on measurement concepts and skills and your students obtained the scores shown in this class list. (Show Table 1) You know that students from this school have had trouble with measurement items on the state test in previous years, and you're wondering whether you need to do more teaching in this area or can move on to the next topic. You take these scores into the teachers' lounge and ask colleagues to take a look. When they ask about the test you explain that you designed it so that if a student gets a score of 80% or better on it, you are really quite confident that he or she understands the concepts. When a student's score is lower than that, you feel there is something they still don't understand.

One of your colleagues pulls out his calculator and shows that the mean for these scores is 80.6. "The mean score is greater than 80. You've done your job. Move on! There's lots more math to cover."

Table 1. Classroom Data Set

Student	Total Score* (% correct)
Aaron	96
Anna	72
Beatrice	92
Bennie	68
Caitlin	92
Chantal	68
Crystal	100
Denny	88
Jaimie	68
Kayti	84
Mickey	68
Noah	96
Patricia	60
Robbie	72
Sofia	84
Stuart	68
Teresa	76
Tyler	68
Victor	100
Zoe	92
Class Mean	**80.6**

* Percentage of test items the student answered correctly.

1.1 Do you agree with this colleague? Why or why not?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Considers score distributions	Mentions that teachers can't rely on the class mean. Points out that about half of the class didn't pass 80. Counts the number of students whose scores are below 80.
Appreciates impact of extreme scores on the mean	Comments that the two students scoring 100 may have pulled up the class mean.
Uses subscale and item data	Comments that would like to see students' test performance broken down by instructional objective or standard before making a decision.

IF teacher says must go on because of pacing calendar, PROBE with: What would you do if there was not this pacing requirement? *If teacher would still proceed, THEN:* Would you have any concerns about this pattern of performance? How do you feel about your students' performance?

1.2 What do you think would happen if you gave the same class of students another test on measurement the next day? What if you gave them the same test again the next day?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands concept of measurement error and variability	Says that results would not necessarily be the same or that some students will likely change "proficiency status."

1.3 Suppose that your students' performance on the various portions of the examination broke down as shown here (Table 2). If you were the teacher, what would you do?
(Show Table 2)

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Examines class means for the four subscales and draws inferences for what needs additional teaching or to be better taught the next time around.
Understands concept of differentiating instruction based on data	Ideally, describes using the assessment results to figure out which students need help in a particular area and then grouping those students and reteaching the subset of skills they need to work on. Alternatively, might describe reteaching all the concepts to students whose scores are below 80% or reteaching concepts of volume and perimeter and area to the whole class.

Table 2. Classroom Data Set by Skill

Student	Length* (% correct)	Weight* (% correct)	Volume* (% correct)	Perimeter and Area* (% correct)	Total Score* (% correct)
Aaron	100	100	100	86	96
Anna	83	67	67	71	72
Beatrice	100	83	83	100	92
Bennie	83	100	67	29	68
Caitlin	100	83	100	86	92
Chantal	67	100	67	43	68
Crystal	100	100	100	100	100
Denny	100	100	83	71	88
Jaimie	83	83	50	57	68
Kayti	100	100	83	57	84
Mickey	83	83	67	43	68
Noah	100	100	100	86	96
Patricia	83	83	50	29	60
Robbie	100	83	67	43	72
Sofia	100	100	83	57	84
Stuart	83	67	67	57	68
Teresa	100	100	67	43	76
Tyler	83	83	67	43	68
Victor	100	100	100	100	100
Zoe	100	100	83	86	92
Class Mean	**92**	**91**	**78**	**64**	**80.6**

* Percentage of test items the student answered correctly.

1.4 What other information/data would you like to get to help you plan your instruction and improve your students' measurement skills?

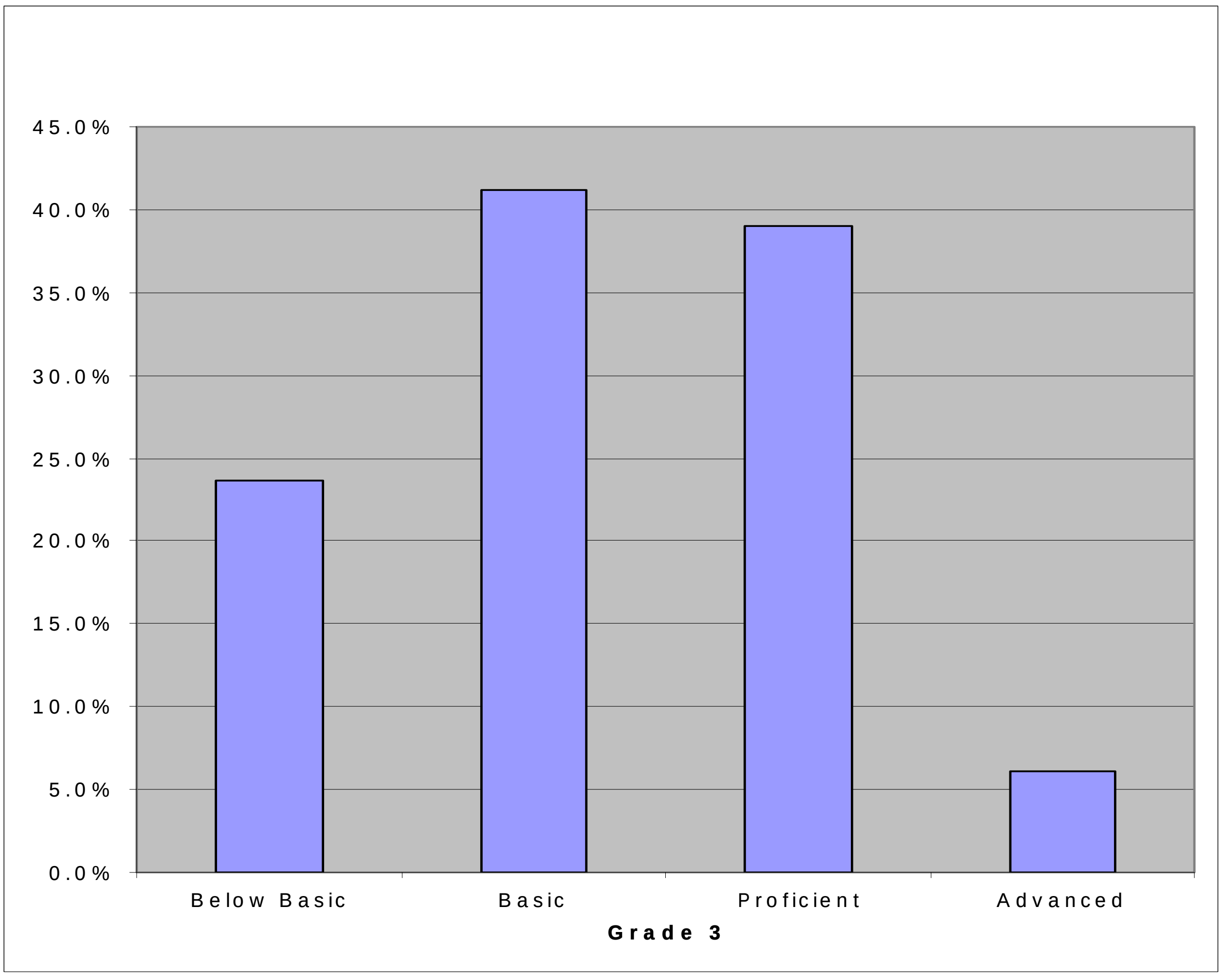
RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Indicates the desire for item-level information. For example, the teacher wants to know: "Which question did a particular student answer wrong?" Indicates the need to understand why a student made certain mistakes. For example, the teacher wants to know "For a particular question, which choice (a, b, c or d) did the student make?" Indicates the need to know which questions are connected to a particular standard or objective.
Appreciates value of multiple measures	Indicates desire for some other, nonmultiple choice assessment, which provides different information (e.g., student classroom work or other quizzes)

SCENARIO 2

SCENARIO: Now I'm going to show you a kind of data display that is a fairly common way to look at how a group of students breaks down in terms of proficiency levels. I'm going to ask you to find some information on the display and then tell me how easy or hard that was to do on a scale from 1 to 10 with 10 being "extremely difficult." Suppose you're in a meeting to discuss 2005–06 reading data from the state assessment for your school's third grade and they hand out this data display at a grade-level meeting. *(Show Figure 1.)*

Figure 1. Reading Achievement Scores Histogram



2.1 Based on this chart, what percentage of the school's third-graders were Advanced in reading?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Says "6 percent" or "a little over 5 percent."

2.2 Based on this chart, what percentage of the school's third-graders were less than proficient in reading?

RESPONSE ANALYSIS

SKILL OR CONCEPT	Evidence for Presence
Finds relevant data in a complex graph	Mentions Basic and Below Basic or the values 41 and 24.
Manipulates data from a complex graph to support reasoning	Says "about 65 percent."

2.3 One of your colleagues, after looking at these data says, "There's something wrong with this chart." Would you agree? Why or why not?

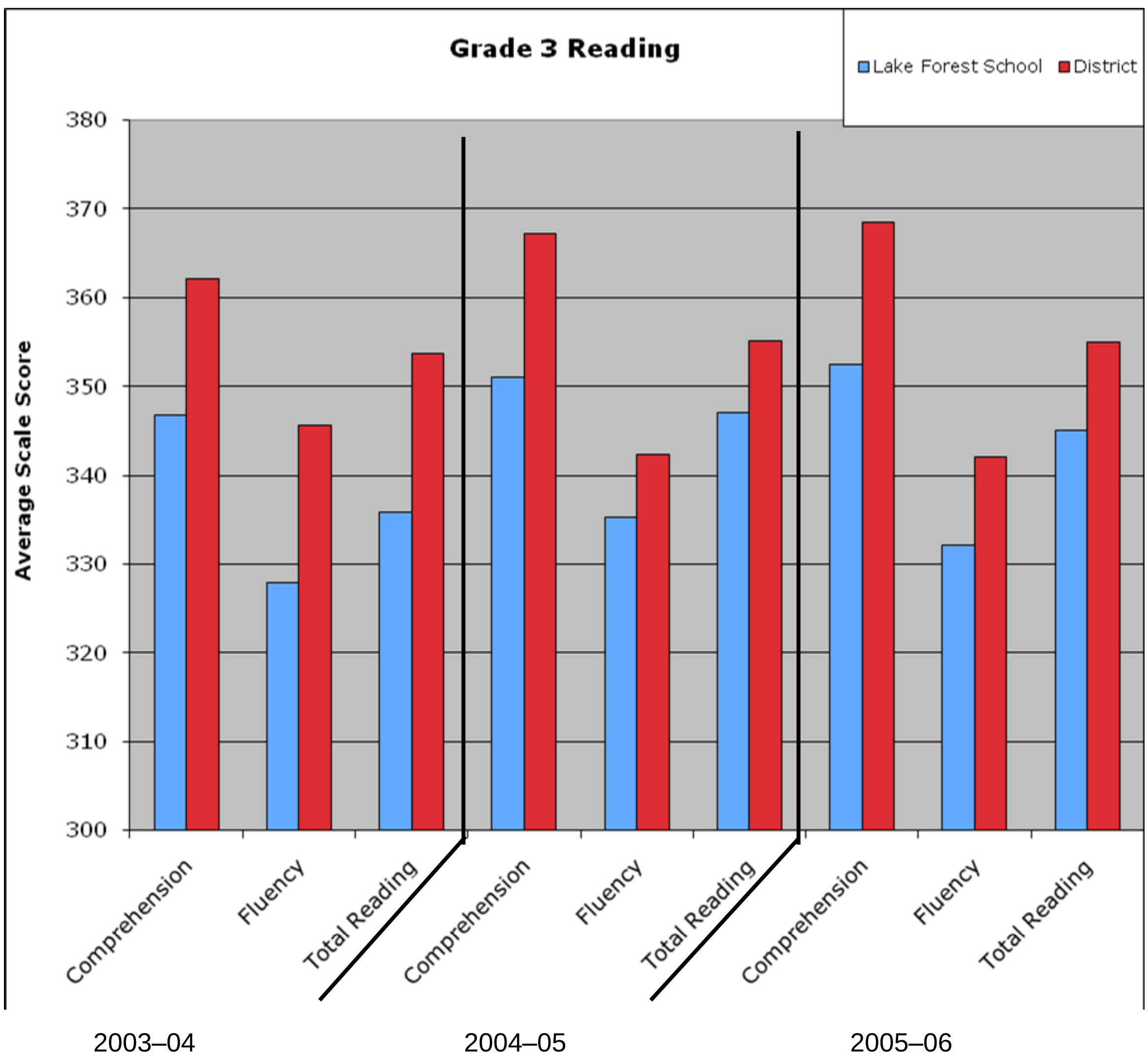
RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Distinguishes between a histogram and a bar chart	Agrees and points out that the depicted percentages add to more than 100 percent.
Manipulates data from a complex graph to support reasoning	Adds the values for the four proficiency categories.

SCENARIO 3

SCENARIO: I'm going to show you a data display of the kind you might have if you are comparing your school to your district. This is a bar graph of Grade 3 reading achievement separated into two components (fluency and comprehension) as well as their total score, for Lake Forest School and its district for each of 3 years. *(Show Figure 2.)*

Figure 2. Grade 3 Reading Achievement Scores Over Three Years



3.1 What was Lake Forest School's average Total Reading Score in 2003–04?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Says "336" or "about 335."

3.2 What was the difference in the district's total reading score in 2005–06 compared to 2003–04?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the values 353 and 355.
Manipulates data from a complex graph to support reasoning	Says "a point or two higher" or "about the same."

Now I'm going to ask you some questions about how you would interpret the data in this chart.

3.3 Looking at the chart as a whole, what would this data tell you about third-graders' reading achievement at this school? (*Get open response.*)

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the 2003–04 and 2005–06 scores; compares mean score from one year to the corresponding score from another year (i.e., Comprehension to Comprehension and Total Score to Total Score).
Manipulates data from a complex graph to support reasoning	Computes score differences as a basis for responding.
Moves fluently between different representations of data	Gives a verbal response consistent with the data manipulations.

OK. Now I'm going to read a series of statements that people might make about the data in this graph and I'd like you to tell me for each one whether you agree or disagree and the reasons why.

a. Although scores fluctuate from year to year, overall Lake Forest School has made improvement in Grade 3 **total reading** since 2003–04.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the values 336 and 345.
Manipulates data from a complex graph to support reasoning	Subtracts the 2003–04 score of 336 from the 2005–06 score of 345.
Moves fluently between different representations of data	Agrees.

b. Compared to the district, Lake Forest School third-graders have made progress in their **total reading** over this three-year period.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the values 336 and 353 for 2003–04 and the values 345 and 355 for 2005–06.
Manipulates data from a complex graph to support reasoning	Computes the gap between the district and the school (353–336) for 2003–04 and the gap between the district and the school (355–345) for 2005–06 in total reading scores. Compares the two differences (17 versus 10).

c. Compared to the district, Lake Forest School third-graders have been making progress in their reading **comprehension** skills over the three-year period.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the values 347 and 352 for the school and the values 362 and 368 for the district.
Manipulates data from a complex graph to support reasoning	Computes the gap between the school and the district (347–362) for 2003–04 and the gap between the school and the district (352–368) for 2005–06 in comprehension scores. Compares the two differences (15 points below the district in 2003–04 and 16 points behind the district in 2005–06).
Moves fluently between different representations of data	Disagrees with the statement as a representation of the data.

d. Lake Forest School's progress in narrowing Grade 3 reading achievement gap compared with the rest of the district has been in reading **fluency** rather than reading **comprehension**.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	For fluency, mentions the values 328 (school) versus 346 (district) for 2003–04, and 332 (school) versus 342 (district) for 2005–06. For comprehension, mentions the values 347 (school) versus 362 (district) for 2003–04 and 352 (school) versus 368 (district) for 2005–06.
Moves fluently between different representations of data	Agrees that statement accurately represents data in the graph. May note that the relative weighting of fluency and comprehension in total reading scores is unknown, but given that the gaps for total reading and for fluency closed while those for comprehension stayed the same or increased slightly, can agree with the statement.

3.4 Suppose Lake Forest School had started using a new reading program at the beginning of the 2004–05 school year while the rest of the district continued with the old program. The annual reading achievement test is given toward the end of academic year. Looking at these data, what are your thoughts about the new curriculum? (*Get open response.*) Which of these statements would you agree with and why?

a. Lake Forest School has not benefited from the new reading program since it was first implemented in 2004–2005.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Examines school values for the three years (335, 348, and 345) in total reading.
Manipulates data from a complex graph to support reasoning	Notes that although the score went down in 2005–06 compared to 2004–05, the school still performed better in reading in 2005–06 compared to how they did in 2003–04. Disagrees with statement.
Understands concept of measurement error and variability	Comments that a small variation such as that between 348 and 345 may not be meaningful.

b. Compared to the old program, the new reading program appeared to help students with their reading comprehension skills.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Mentions the values 347, 351, and 353 for the school and 362, 367, and 368 for the district.
Manipulates data from a complex graph to support reasoning	Compares increase for the school and district. Disagrees since the gap between the school and the district in reading comprehension has remained the same (five to six points difference every year).
Uses subscale and item data	Says would like to see performance by test item and to compare the test items to the new curriculum.

c. Scores move around from year to year, but the new reading program appears promising for enhancing students' reading fluency.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex graph	Examines school fluency subtest values for the three years (328, 335, and 332).
Manipulates data from a complex graph to support reasoning	Notes that although the score went down in 2005–06 compared to 2004–05, the school still performed better in fluency in 2005–06 than they did in 2003–04.
Understands concept of measurement error and variability	Comments that small variations are expected from year to year due to chance. School subtest scores are likely to be less stable than those of districts because there are fewer examinees within a single school than within the district.
Appreciates value of multiple measures	Indicates desire to collect additional years of data or to look at data from additional measures of reading fluency.
Uses subscale and item data	Says would like to see performance by test item and to compare the test items to the new curriculum.

3.5

a. Are there other possible explanations, other than the effect of the new reading program, that might explain the pattern of results for Lake Forest third-graders over these three years? *(Get open response.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Distinguishes between cross-sectional and longitudinal data	Points out that the three years of data are for different student groups and these are not necessarily comparable from year to year.
Appreciates value of multiple measures	Says would like to get other reading assessment results for the school and district for the same time period.

b. Would you agree or disagree with the following statement and why:

You can’t be sure whether the program is having an effect because each year different third-graders take the reading achievement test.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Distinguishes between cross-sectional and longitudinal data	Agrees. Points out that the three years of data are for different student groups and these are not necessarily comparable from year to year.
Appreciates value of multiple measures	Says would like to get other reading assessment results for the school and district for the same time period.

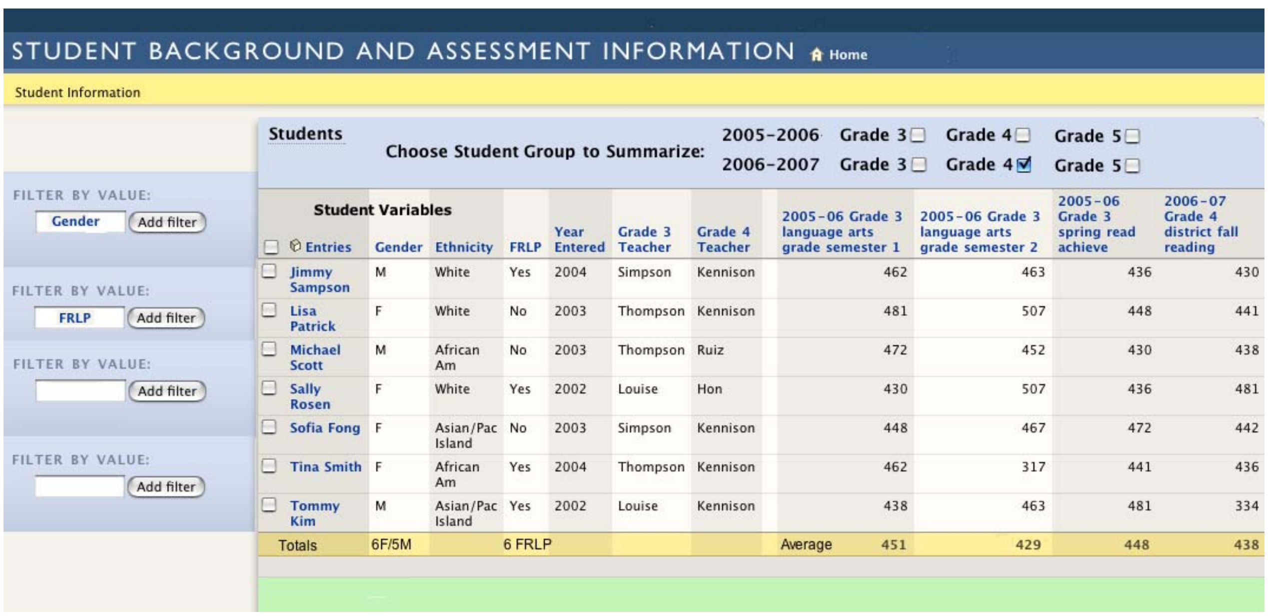
Scenario 4

SCENARIO: Now I'm going to describe a hypothetical situation and a computer-based student data system to you and I'd like to see what kind of information you'd like to get from the system.

Suppose it's January 2008 and you're one of the fourth-grade teachers in a school that was surprised by fourth graders' relatively low performance on the state reading test last year (spring 2007). Your principal has encouraged you to use student data to gain insights into how you can get higher Grade 4 achievement this year.

The data system available to you (see Figure 3) contains data on both current (2007–08) fourth graders and last year's (2006–07) third graders (see Student Groups in Table 3) as well as other student groups. For each student, the data system has (1) scores on last spring's state reading test, (2) scores on a district test given in the fall (also shown in Table 3), and (3) semester grades in language arts. It also has other information about students that can be used to create subgroups within a grade if you want to see how different subgroups compare–for example, ethnicity, gender, and whether the student is eligible for free or reduced-price lunch (FRPL).

Figure 3. Screenshot From the Data System



Entries	Gender	Ethnicity	FRLP	Year Entered	Grade 3 Teacher	Grade 4 Teacher	2005-06 Grade 3 language arts grade semester 1	2005-06 Grade 3 language arts grade semester 2	2005-06 Grade 3 spring read achieve	2006-07 Grade 4 district fall reading
Jimmy Sampson	M	White	Yes	2004	Simpson	Kennison	462	463	436	430
Lisa Patrick	F	White	No	2003	Thompson	Kennison	481	507	448	441
Michael Scott	M	African Am	No	2003	Thompson	Ruiz	472	452	430	438
Sally Rosen	F	White	Yes	2002	Louise	Hon	430	507	436	481
Sofia Fong	F	Asian/Pac Island	No	2003	Simpson	Kennison	448	467	472	442
Tina Smith	F	African Am	Yes	2004	Thompson	Kennison	462	317	441	436
Tommy Kim	M	Asian/Pac Island	Yes	2002	Louise	Kennison	438	463	481	334
Totals	6F/5M		6 FRLP			Average	451	429	448	438

Table 3. Student Data Available in the System

Student Background Variables

- Ethnicity
- Gender
- Free or reduced-price lunch
- Year entered district
- Grade 3 teacher
- Grade 4 teacher

Student Achievement Data

Spring 2007 State Assessment (Score Range: 0–650)—last school year

- Spring 2007 Grade 3 state reading achievement score
- Spring 2007 Grade 4 state reading achievement score
- Spring 2007 Grade 5 state reading achievement score

Fall 2007 District Assessment (Score Range: 0–50)—current school year

- Fall 2007 Grade 3 district reading test
- Fall 2007 Grade 4 district reading test
- Fall 2007 Grade 5 district reading test

2006–07 Language Arts Semester Grades
(O= Outstanding; S= Satisfactory; I= Improving; U= Unsatisfactory)—last school year

- Grade 3 semester 1 (2006–07)
- Grade 4 semester 1 (2006–07)
- Grade 5 semester 1 (2006–07)
- Grade 3 semester 2 (2006–07)
- Grade 4 semester 2 (2006–07)
- Grade 5 semester 2 (2006–07)

Student Groups in the Data System

- 3rd graders (2007–08)
- 3rd graders (2006–07)
- 4th graders (2007–08)
- 4th graders (2006–07)
- 5th graders (2007–08)
- 5th graders (2006–07)

4.1 So now in January 2008, what specific achievement data would you want to get from this system to help you decide how to improve your fourth-graders' reading performance? (*Follow-up probes*:) Tell me what group of students and which testing period you want to focus on.

a. Why do you want to access that data and how do you plan to use the data once you get it?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Aligns question with purpose and data	Picks a logical group (e.g., either 2006–07 third-graders or 2007–08 fourth-graders) AND selects a logical measure for that group (e.g., either spring 2007 Grade 3 state test scores or fall 2007 Grade 4 district fall test scores).

4.2 Would you like to make any other queries of the data system? *(Follow-up probe:)* Are there any other questions you want to ask that can be answered by the data in the system?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Forms queries that lead to actionable data	Wants to find out whether student performance is related to their previous teachers. Expresses a desire to identify those teachers whose students show the greatest gains so that whole grade can learn from their practices.
Aligns question with purpose and data	Picks a logical set of filters and outcome measure given their question.
Appreciates value of multiple measures	Follows up a query about one student learning outcome with a comparable query using another measure. Wants to compare performance on the prior spring's state test with fall test performance of the same students.

4.3 To improve this year's fourth-graders' achievement, are there other data you would like to have that you don't see represented in this system?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Forms queries that lead to actionable data	Wants to investigate relationship between student performance and receipt of special services such as tutoring. Expresses a desire to identify those teachers whose students show the greatest gains so that whole grade can learn from their practices.
Uses subscale and item data	Says would like to see performance by test item or content standard.
Appreciates value of multiple measures	Expresses desire for additional measures of student learning. Wants to compare performance on the prior spring's state test with fall test performance of the same students.

4.4 Can you think of any investigations you might want to conduct to help you understand your students and improve their achievement?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Aligns question with purpose and data	States a question and describes a kind of data and an analysis that would address that question.
Forms queries that lead to actionable data	Wants to investigate relationship between student performance and variables under school or teacher control. Expresses a desire to identify those teachers whose students show the greatest gains so that whole grade can learn from their practices.
Uses subscale and item data	Says would like to see performance by test item or content standard.
Appreciates value of multiple measures	Expresses desire for additional measures of student learning (e.g., report cards). Wants to compare performance on the prior spring's state test with fall test performance of the same students.

SCENARIO 5

SCENARIO: This is the kind of data table (Table 4) that some student data systems produce. We're going to do a warm up by finding some information from the table. (Show Table 4.)

Table 4. 2006–07 Score Levels—Mathematics

Grade	Gender	Ethnicity	Number of Students Tested	Percent of Tested Students	Mean Scale Score	Number Students at Each Proficiency Level			
						Below Basic	Basic	Proficient	Advanced
3	Female	African American	1	1%	589	0	0	0	1
		Asian/Pac Islander	18	26%	444	5	4	6	3
		Latino	17	24%	428	6	5	5	1
		White	34	49%	449	4	12	12	6
		Total Female	70	100%	445	15	21	23	11
	Male	African American	2	3%	452	0	1	0	1
		Asian/Pac Islander	18	23%	450	3	6	6	3
		Latino	31	40%	430	8	7	14	2
		White	27	35%	448	6	11	7	3
		Total Male	78	100%	440	17	25	27	9
4	Female	African American	2	3%	462	1	0	1	0
		Asian/Pac Islander	20	26%	472	2	7	8	3
		Latino	18	24%	441	3	8	5	2
		White	36	47%	436	8	12	12	4
		Total Female	76	100%	447	14	27	26	9
	Male	African American	0	0%	NA	0	0	0	0
		Asian/Pac Islander	16	23%	442	2	8	5	1
		Latino	24	35%	438	3	13	5	3
		White	29	42%	456	5	12	10	2
		Total Male	69	100%	446	10	33	20	6
5	Female	African American	1	1%	317	1	0	0	0
		Asian/Pac Islander	35	32%	470	6	6	8	6
		Latino	22	29%	452	4	7	8	3
		White	22	37%	470	5	8	10	5
		Total Female	80	100%	463	14	21	26	14
	Male	African American	3	4%	560	0	0	1	2
		Asian/Pac Islander	18	26%	458	4	5	5	4
		Latino	16	24%	449	2	5	6	3
		White	31	46%	464	4	12	13	2
		Total Male	68	100%	462	10	22	25	11

5.1 What was the mean (or average) scale score for the Asian/Pacific Islander fourth-grade girls who took the test?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Identifies 472 as the mean score.

5.2 Which student group had the lowest average or mean mathematics scale score in Grade 4?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Examines Grade 4 student subgroup mean scale scores: 462, 472, 441, 436, NA, 442, 438, and 456.
Manipulates data from a complex table to support reasoning	Compares various subgroup mean scale scores. Answers white females.

5.3 How many Asian/Pacific Islander fourth-grade boys took the test?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Identifies 16 as the number of Asian/Pacific Islander fourth-grade boys.

Now I'd like you to look at the data in this table for Grade 3.

5.4 Overall, based on the Grade 3 data in this table, would you say that there was a difference between boys and girls in mathematics test performance?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Considers score distributions	Looks at proficiency category distributions, not just mean scale scores.
Manipulates data from a complex table to support reasoning	Computes proportion of males and proportion of females in a proficiency category using the provided frequencies.
Interprets a contingency table	Points out that it is difficult to compare boys' and girls' performance because there is a very different distribution of ethnic groups for the two genders.
Appreciates impact of extreme scores on the mean	Points out the extremely high score of the African-American third-grade girl as likely pulling up the girls' mean.
Understands concept of measurement error and variability	The difference between boys' and girls' mean scale scores is quite small; it may not indicate any difference in "true" scores.

5.5 Now I'd like you to think about the implications of the Grade 3 data. Remember that these are for last year's third-graders in the 2006–07 school year. If there have been no major changes in the school's student body, teachers, or curriculum, would you expect on the basis of these data that:

a. This year's third-grade girls can be expected to score higher than boys when this test is given to this year's third-graders.

RESPONSE Analysis

Skill or Concept	Evidence for Presence
Understands concept of measurement error and variability	The difference between boys' and girls' mean scale scores this year is quite small; it may not indicate any difference in "true" scores so we wouldn't necessarily expect it to be repeated next year.
Considers score distributions	Disagrees. The proportion of boys and girls attaining proficiency this year was equal.
Appreciates impact of extreme scores on the mean	Points out the extremely high score of the African-American third-grade girl as likely pulling up the girls' mean scale score.
Understands relationship between sample size and generalizability	Comments that the extremely high score contributed by one girl would not necessarily be seen in the next year. Points out that little can be said concerning gender differences for African-American students because there were so few of them.
Interprets a contingency table	Points out that boys and girls had roughly equal mean scale scores if you look within ethnic group (except for African-Americans).

b. This year's third-grade African-American girls will score better than other students when this test is given to this year's third-graders.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands relationship between sample size and generalizability	Points out that little can be said concerning likely performance of next year's African-American girls based on this year's sample of only one.

5.6 Now let's assume that you're a third-grade teacher and these Grade 3 data are for mid-year performance on a benchmark test. Are there particular students you think will be most likely to have trouble scoring Basic or above on the state test at the end of the year? (*If appropriate, probe with* <u>*one*</u> *of the following.*) Which students would you be concerned about and what data trigger that concern? *OR* Why don't you think the data in the table point to a need for intensive support for any particular students?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Says the students who are "Below Basic" rather than "Latino" or "Latino girls."
Considers score distributions	Understands that a subgroup mean does not tell you where each individual in the subgroup scored. Bases response on data about students' performance level rather than on ethnicity or gender label.

Finally, let's look at the data for Grade 5.

5.7 I'm going to read a series of statements that people might make about different aspects of the Grade 5 data in this table. I'd like you to tell me for each statement whether you agree or disagree and the reasons why.

a. A majority of fifth-graders at this school have achieved proficiency in mathematics as measured by this test.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Attends to number of Proficient and Advanced students within each gender (26, 14, 25, 11).
Manipulates data from a complex table to support reasoning	Sums the number of Proficient and Advanced girls and boys and divides by the total number of fifth-graders to get 51 percent achieving proficiency.
Moves fluently between different representations of data	Agrees that the statement accurately represents data in the table.

b. In Grade 5, girls were more likely than boys to score Below Basic on this assessment.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Attends to number of Below Basic girls (14) and total girls (80) and the number of Below Basic boys (10) and total boys (68).
Manipulates data from a complex table to support reasoning	Computes the relevant percentages: 16 percent of girls and 19 percent of boys.
Moves fluently between different representations of data	Disagrees. Statement is not an accurate representation of the data in the table.

c. Of those students who scored Advanced in Grade 5, most were Asian/Pacific Islanders.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Examines number of students categorized as Advanced (14 girls and 11 boys) and the number of Advanced Asian/Pacific Islanders (six girls and four boys).
Manipulates data from a complex table to support reasoning	Computes the percentage: 10 out of 25 or 40 percent, which is less than half of the students scoring in the Advanced range.
Moves fluently between different representations of data	Disagrees. Statement is not an accurate representation of the data in the table. Less than half cannot be “most.”

SCENARIO 6

SCENARIO: Suppose you're teaching in a district that requires students to attain eighth-grade proficiency in mathematics in order to enroll in Algebra I in high school. Looking at students' performance the preceding year, you found the results in this table for the Latino and African American students who make up your school's entire student body. (Show Table 5.)

A score of 35 on the district math test is considered proficient, and

A school is considered "low performing" if less than 50 percent of students in one of the student subgroups reach proficiency.

Table 5. Achievement in Grade 8 Mathematics

Group	Number of Students	Group Mean Math Score	Number of Students Proficient	Percent Proficient
Latino	239	38.5	143	60%
African American	52	36.5	25	48%

6.1 What do these data tell you about how well students are doing at your school? *(Get open response first.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Manipulates data from a complex table to support reasoning	Combines data across the two student subgroups.
Moves fluently between different data representations	Makes an accurate verbal statement about the data (for example, that about 168 of the school's 291 students are proficient).

Now I'm going to read you some statements again and I'd like you to tell me whether you agree or disagree with each and why. (*For each statement, give teacher an index card with the statement on it to keep handy as he or she looks at the data.*)

a. Both of our student groups are doing pretty well in math since their mean scores are above 35.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Considers score distributions	Points out that not every student has a score near the mean, and given the fairly low percent proficient in the two subgroups, there must be many students who have not yet attained proficiency.

b. More than half of our eighth-graders are proficient in eighth-grade math.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Manipulates data from a complex table to support reasoning	Applies the percent proficient to the number of students in each subgroup to estimate that about 168 of the school's 291 students are proficient, which is well more than half.
Moves fluently between different representations of data	Agrees that the statement accurately reflects data in the table.

c. Our school is not getting enough African-American students to proficiency, but Latino students are meeting the required performance standard as a subgroup.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant data in a complex table	Examines the percentage of Latino students proficient (60 percent) and the percentage of African-American students proficient (48 percent) and compares both to the district requirement of 50 percent or more.
Moves fluently between different representations of data	Agrees that the statement accurately reflects data in the table.

d. Our school is classified as low-performing based on these mathematics scores.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Finds relevant cells in a complex table	Compares the percentage of African-American students proficient (48 percent) to the district's standard (at least 50 percent of every subgroup proficient).
Moves fluently between different representations of data	Agrees that the statement accurately reflects the data in the table.

6.2 What actions should your school consider to avoid being labeled "low-performing" in the coming year? What other information or data do you need? *(Get open response.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Appreciates value of multiple measures	Suggests assessing students early and frequently during the year to identify individual students who are struggling.
Uses subscale and item data	Suggests looking at particular items or skills in which students had trouble as a guide to what to emphasize in instruction.
Understands concept of differentiating instruction based on data	Suggests grouping students for instruction or intensive support based on their performance on interim assessments, NOT based on subgroup identification.
Understands relationship between sample size and generalizability	Notes that the school does not have a large number of African-American students so the scores of the new cohort of African-Americans may well be different from the prior year's results.
Understands concept of measurement error	Suggests that next year's results for African-American students could be different even if the students are similar.

6.3 Which of these statements do you agree with based on these data? Explain your answer for each. (*For each statement, give teacher an index card with the statement on it to keep handy as he or she looks at the data.*)

a. This year all African-American students should get supplemental instruction to improve their math performance.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands relationship between sample size and generalizability	Notes that the school does not have a large number of African-American students so the scores of the new cohort of African-Americans may well be different from the prior year's results.
Understands concept of measurement error	Suggests that next year's results for African-American students could be different even if the students are similar.
Understands concept of differentiating instruction based on data	Suggests grouping students for instruction or intensive support based on their performance on interim assessments NOT based on subgroup identification.

b. The school does not need to provide supplemental instruction to Latino students this year since their group met the proficiency criterion last year.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands concept of measurement error	Suggests that next year's results for Latino students could be different even if the students are similar. The Latino students' mean score was not far above the cutoff.
Understands concept of differentiating instruction based on data	Suggests grouping students for instruction or intensive support based on their performance on interim assessments NOT, based on subgroup identification.
Considers score distributions	Although the Latino subgroup met the proficiency criterion, there were many individual Latino students who did not pass the proficiency standard. Students' individual scores should be examined to identify those students who need extra support.

c. If no changes are made to the current eighth-grade math program and next year's students are similar to this year's, there's a reasonable chance that 50 percent or more of next year's eighth-grade African-American students will meet the proficiency requirement.

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands relationship between sample size and generalizability	Notes that the school does not have a large number of African-American students, so the scores of the new cohort of African-Americans may well be different from the prior year's results.
Understands concept of measurement error	Suggests that next year's results for African-American students could be different even if the students are similar.

SCENARIO 7

SCENARIO: Suppose that this is the third week of school and that you're a third-grade teacher planning your instruction for the remainder of this term. As shown here (Table 6), you have scores from the state reading test given last spring and from a sight reading assessment and a passage comprehension test that you've had your students take during the first two weeks of school. (Show Table 6.)

Table 6. Student Performance on State and Classroom Reading Tests

Student	2006–07 State Achievement Test Scale Score				Fall 2007 Class Tests	
	Total Reading	Vocabulary	Comprehension		Sight Reading	Text Comprehension
Aaron	393	375	410		16	5
Anna	530	510	550		24	7
Beatrice	498	505	490		22	8
Bennie	528	515	540		26	9
Caitlin	645	660	630		28	12
Chantal	513	515	510		20	10
Crystal	573	560	585		24	10
Denny	588	566	610		20	6
Jaimie	555	550	560		25	10
Kayti	541	553	528		26	9
Mickey	410	395	425		16	5
Noah	693	678	700		30	11
Patricia	416	400	432		20	7
Robbie	563	580	545		26	8
Sofia	480	500	460		22	10
Total Possible	700	700	700		30	12
Class Average	530	527	532		23	8

7.1

a. What data would you look at as you're planning your instruction? Which data would be most important to you and why? *(Get open response.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands value of multiple measures	Notes the advantage of looking at more than one assessment. Statewide tests usually have stronger technical quality (reliability) but may not match the local curriculum and may not include content above Grade 2. Also, the classroom tests were given more recently.

b. What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class? *(Get open response.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Discusses looking at comprehension performance, sight reading (fluency), and vocabulary scores separately to decide what to emphasize with each group and within each group, what each student is struggling with, and how to accommodate their needs.
Understands concept of differentiating instruction based on data	Articulates a coherent, data-informed rule for grouping (i.e., includes test data but data need not be the sole criterion).

c. Are there other kinds of information you would like to have to support your instructional planning? *(Get open response.)*

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands value of multiple measures	Says would ask for other student information, such as: ... Second-grade report card, indicating reading at, above or below grade level ... Other benchmark or formative reading assessments conducted in second grade ... Other standardized assessment results ... Second-grade teachers' notes

7.2

a. Would you place students in different small reading groups for this instruction?

If YES: How would you group students and how would your instruction vary for the different groups?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Understands concept of differentiating instruction based on data	Articulates different content or pedagogy for the groups or for individual students consistent with their score profiles (e.g., assign different books or provide different amounts of direct teaching or guided reading to different groups).

7.3

a. Which group would you put Aaron into? *OR* Which approach would you use for Aaron? Why?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Discusses comprehension performance, sight reading (fluency), and vocabulary scores separately to decide what to emphasize with each group. Indicates a desire to look more closely at Aaron's performance on different items or standards to support a diagnosis of his needs.
Appreciates value of multiple measures	Discusses comprehension scores on both the state test and the classroom test. Indicates a desire to obtain additional information, such as a reading specialist's assessment of Aaron or notes from his second-grade teacher.
Understands concept of differentiating instruction based on data	Discusses an individualized instructional plan for Aaron based on his assessment results.

b. Which group would you put Denny into? What is your reason for that decision?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Notes the discrepancy between Denny's high scores on the state test and his below-average scores on the in-class tests of reading comprehension and sight reading. Indicates desire to have more detailed information about Denny's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.
Understands the concept of measurement error and variability	Notes that Denny's score on the in-class test of reading comprehension could be an aberrant result.
Appreciates value of multiple measures	Suggests getting another assessment of Denny's reading comprehension and sight reading, either formally or through in-class observation. Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.

c. Which group would you put Sofia into? Why?

RESPONSE ANALYSIS

Skill or Concept	Evidence for Presence
Uses subscale and item data	Notes the discrepancy between Sofia's relatively low scores on the state test last spring and her strong performance on the in-class tests this fall. Indicates desire to have more detailed information about Sofia's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.
Understands the concept of measurement error and variability	Notes that Sofia's state test performance last fall might underrepresent her skills.
Appreciates value of multiple measures	Suggests getting additional information about Sofia's reading ability (e.g., report card, second-grade end of semester reading test). Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.

References

Adelman N., M. B. Donnelly, T. Dove, J. Tiffany-Morales, A. Wayne, and A. Zucker. (2002). *The integrated studies of educational technology: Professional development and teachers' use of technology.* Arlington, Va.: SRI International.

Alloy, L. B., and N. Tabachnik. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, 91(1), 1120149.

American Recovery and Reinvestment Act of 2009, S. 1, 111th Congress, 1st Session. (2009).

Batanero, C., A. Estepa, J. D. Godino, and D. R. Green. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education,* 27: 1510169.

Besterfield-Sacre, M., and E. R. Halverson. (2007). *Using systems engineering approaches to improve urban school districts: Process mapping and process measurement.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Ill.

Birkeland, S., E. Murphy-Graham, and C. H. Weiss. (2005). Good reasons for ignoring good evaluation: The case of the drug abuse resistance education (D.A.R.E.) program. *Evaluation and Program Planning,* 28: 247–256.

Bright, G. W., and S. N. Friel. (1998). Graphical representations: Helping students interpret data. *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 63088). Mahwah, NJ: Erlbaum.

Chen, E., M. Heritage, and J. Lee. (2005). Identifying and monitoring students' learning needs with technology. *Journal of Education for Students Placed At Risk*, 10(3): 3090332.

Choppin, J. (2002). *Data use in practice: Examples from the school level.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Coburn, C. E., M. I. Honig, and M. K. Stein. (forthcoming). What is the evidence on districts' use of evidence? In *Research and practice: Towards a reconciliation,* eds. J. Bransford, L. Gomez, D. Lam, and N. Vye, Cambridge, Mass.: Harvard Educational Press.

Coburn, C. E., J. Toure, and M. Yamashita. (forthcoming). Evidence, interpretation, and persuasion: Instructional decision making at the district central office. *Teachers College Record*.

Confrey, J., and K. M. Makar. (2005). Critiquing and improving the use of data from high-stakes tests with the aid of dynamic statistics software. In *Scaling up success: Lessons learned from technology-based educational improvement,* eds. C. Dede, J. P. Honan, and L. C. Peters, 198–226. San Francisco, Calif.: Jossey-Bass.

Consortium for School Networking. (2004). *Vision to know and do: The power of data as a tool in educational decision making.* Washington, D.C.: Author.

Curcio, F. R. (1989). *Developing graph comprehension: Elementary and middle school activities*. Reston, Va.: National Council of Teachers of Mathematics.

Diamond, J. B., and K. Cooper. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. In *Evidence and decision making, 106th Yearbook of the National Society for the Study of Education,* ed. P. A. Moss, 241–263. Malden, Mass.: Blackwell.

Elementary and Secondary Education Act of 2001, H.R. 1, 107th Congress. (2002).

Elstein, A. S., L. S. Shulman, and S. A. Sprafka. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, Mass.: Harvard University Press.

Feldman, J., and R. Tung. (2001). Using data based inquiry and decision-making to improve instruction. *ERS Spectrum,* 19(3): 10–19.

Hamilton, L., R. Halverson, S. Jackson, E. Mandinach, J. Supovitz, and J. Wayman. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Hertwig, R., and G. Gigerenzer. (1999). The conjunctive fallacy revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making,* 12: 275–305.

Hoffman, L. (2007). *Numbers and type of public elementary and secondary education agencies for the Common Core of Data: School Year 2005–06 (NCES 2007-353),* U.S. Department of Education. Washington, D.C.: National Center for Education Statistics.

Honig, M. I. (2003). Building policy from practice: District central office administrators' roles and capacity for implementing collaborative education policy. *Educational Administration Quarterly,* 39(3): 2920338.

Huang, W., P. Eades, and S-H. Hong. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization,* 8, 1390152.

Kennedy, M. M. (1982). Evidence and decision. In *Working knowledge and other essays,* ed., M. M. Kennedy, 59–103. Cambridge, Mass.: The Huron Institute.

Khanna, R., D. Trousdale, W. R. Penuel, and J. Kell, (April 1999). *Supporting data use among administrators: Results from a data planning model.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec.

Konold, C. (2002). Alternatives to scatterplots. In *Proceedings of the Sixth International Conference on Teaching Statistics,* ed., B. Phillips. Voorburg, The Netherlands: International Statistics Institute.

Lawless, K. A., and J. W. Pellegrino. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research,* 77(4): 575–614.

Noss, R., Pozzi, S., and C. Hoyles. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics,* 40(1), 25051.

Penuel, W. R., J. Kell, J. Frost, and R. Khanna. (April 1998). *Administrator reasoning about data.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, Calif.

Porter, A. C., M .S. Garet, L. Desimone, K. S. Yoon, and B. F. Birman. (2000). *Does professional development change teaching practice? Results from a three-year study* (U.S. Department of Education Report No. 2000004). Washington, D.C.: U.S. Department of Education.

Ranney, M. A., L. F. Rinne, L. Yarnall, E. Munnich, L. Miratrix, and P. Schank. (2008). Designing and assessing numeracy training for journalists: Toward improving quantitative reasoning among media consumers. In *International perspectives in the learning sciences: Proceedings of the eighth International Conference for the Learning Sciences* (Vol. 2) , eds. P. A. Kirschner, F. Prins, V. Jonker and G. Kanselaar, pp. 2460253. Utrecht, the Netherlands: ISLS.

Schmoker, M. J. (1996). *Results: The key to continuous school improvement.* Alexandria, Va.: Association for Supervision and Curriculum Development.

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, 18(2): 141–179.

Thorn, C. A. (2002). *Data use in the classroom: The challenges of implementing data-based decision-making at the school level.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Tversky, A., and D. Kahneman. (1982). Judgment under uncertainty: Heuristics and biases. In *Judgment under uncertainty: Heuristics and biases,* eds., D. Kahneman, P. Slovic, and A. Tversky, 3–20. New York: Cambridge University Press.

U.S. Department of Education, Office of Educational Technology. (2004). *Toward a new golden age in American education: How the Internet, the law and today's students are revolutionizing expectations.* Washington, D.C.: Author.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007*. Washington, D.C.: Author.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2009). *Implementing data-informed decision making in schools: Teacher access, supports, and use*. Washington, D.C.: Author.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *Use of education data at the local level: From accountability to instructional improvement*. Washington, D.C.: Author.

Wayman, J. C. and S. Stringfield. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education,* 112(4): 549–571.West, R. F., and C. Rhoton. (1994). School district administrators' perceptions of educational research and barriers to research utilization. *ERS Spectrum,* 12(1): 23–30.

Westbrook, J. I., A. S. Gosling, and E. Coiera. (2004). Do clinicians use online evidence to support patient care? A study of 55,000 clinicians. *Journal of the American Medical Informatics Association,* 11: 113–120.

Westbrook, J. I., A. S. Gosling, and E. Coiera. (2005). The impact of an online evidence system on confidence in decision making in a controlled setting. *Medical Decision Making,* 25: 178–185.

Willis, G. B. (1999). *Cognitive interviewing: A "how to" guide.* Short course presented at the 1999 meeting of the American Statistical Association.