Research Report

An Evidence-Centered Approach to Using Assessment Data for Policymakers

Jody S. Underwood

Diego Zapata-Rivera

Waverely VanWinkle

January 2010

ETS RR-10-03

Listening. Learning. Leading.®

An Evidence-Centered Approach to Using Assessment Data for Policymakers

Jody S. Underwood
Pragmatic Solutions, Westlake Village, California

Diego Zapata-Rivera and Waverely VanWinkle
ETS, Princeton, New Jersey

January 2010

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Technical Review Editor: Don Powers

Technical Reviewers: Irv Katz and Teresa Egan

Abstract

District-level policymakers are challenged to use evidence of student achievement to make policy decisions, such as professional development and other school improvement plans. They currently receive reports of student achievement data that are complex, difficult to read, and even harder to interpret. Using the research literature on policymakers' use of data and conducting focus groups and interviews, we elicited information on their roles and responsibilities, as well as questions these people would like to have answered from achievement data. We propose an evidence-centered reporting framework to help policymakers determine which data they need, in order to design a series of reports that will answer their questions and to help them make sense of the data in support of policy decisions.

Key words: student achievement, professional development, evidence-centered design, ECD, educational policy

Acknowledgments

We would like to thank our Web page designer, Debbie Pisacreta. We also want to acknowledge our colleges Irv Katz, Teresa Egan, and Don Powers for providing insightful comments on a previous version of this paper.

Table of Contents

List of Figures

Page

Introduction

The majority of effort concerning the formative use of assessment data has focused on the classroom and the teacher, with very little focusing on local policymakers. This is a serious oversight (Spillane, 2004, 2005), since local policymakers routinely make critical decisions that have direct effects on what goes on in classrooms.

Policymakers are present at local, state, and federal levels and are involved in decision-making processes that influence a variety of people and practices. It is vital that policymakers understand evidence use at the district level, in addition to the surrounding conditions, in order to provide the specific supports that are needed (Honig & Coburn, 2008). Where score reports containing evidence of student achievement are provided to policymakers, it has been found that they are presented in ways that are not easily interpretable by these stakeholders (Hambleton, 2007; Hambleton & Slater, 1994). In order to understand how policy is implemented at the local level, it is beneficial to look at what district policymakers do and do not do in terms of the ideas they generate and actions they take (Spillane 1998a, 1998b cited in Spillane, 2000).

One of the difficulties that policymakers face is that they often receive conflicting messages, sometimes from the same source, and often from different sources. Thus, Honig and Hatch (2004) have identified the local policymaker's challenge as one of "crafting coherence" from the multiple external demands they experience. Given these inconsistencies, administrators must choose to ignore certain demands, accommodate others, and reinterpret others. Honig and Hatch make very clear that the vision of policy leading directly to practice, particularly in light of conflicting messages, is an idealized fiction. Assuming that administrators always make decisions based only on sound evidence is also unrealistic.

An important research question is whether assessment reports for administrators can serve the same kind of formative role in fostering coherence as has been shown to be the case for teachers using classroom assessments. Many classroom efforts have used assessment results to help teachers connect student outcomes to the actions of their instruction and to refine their instruction as a consequence (e.g., Ball & Cohen, 1999; Pellegrino, Chudowsky, & Glaser, 2001; Shepard, 2000). These efforts all include teachers using assessment data to provide insight into what students know and can do, which has led to instructionally relevant decisions for individuals or groups of students.

Our ultimate goal is to design reports that will help policymakers discover and understand student achievement patterns and to provide interpretive recommendations about the implications of particular results for decisions that are within the scope of their responsibilities. However, before we can undertake this, we first need to articulate a framework for designing the reports and providing access to the right data at the right time and in easily accessible ways.

This paper presents an evidence-based reporting framework for helping district-level policymakers find and use assessment data appropriate for their needs. An example based on the Cognitively Based Assessment of, for, and as Learning (CBAL) project (Bennett & Gitomer, 2009) is used to illustrate this new approach.

Policymaker Use of Assessment Data for Decision Making

What are the instructionally relevant decision-making needs of district-level policymakers, and how can assessment reports be used to support these decisions? Clearly, the level of granularity that a teacher needs for a particular student is not going to be helpful to a district-level administrator. But there are patterns of results that may be helpful to support decisions about curriculum selection, professional development needs of teachers, instructional methods, and so on, as well as the needs of particular groups of students. To achieve our goals, we start with a description of district-level policymaker responsibilities and the types of decisions they make. Then, we describe how policymakers use evidence in decision making.

Responsibilities and Decisions of Policymakers

District central offices play a critical role in improving classroom instruction and raising student achievement in schools through communication with principals, teachers, and students (Mac Iver & Farley, 2003). Through a review of the literature, we have identified seven types of responsibilities that fall under the roles of these policymakers (see Figure 1):

1. School improvement plans (Honig, 2003; Honig & Coburn, 2008; Miller, 2003; Wayman, Midgley, & Stringfield, 2005)
2. Professional development (Brunner et al., 2005; Coburn, Honig, & Stein, 2009; Coburn, Toure, & Yamashita, in press; Honig & Coburn, 2008; Mac Iver & Farley, 2003)

3. Program selection and evaluation (Brunner et al., 2005; Coburn & Talbert, 2005; Guerard, 2001; Honig, 2003; Honig & Coburn, 2008)
4. Curriculum selection (Coburn et al, 2009; Coburn et al., in press; Honig & Coburn, 2008; Mac Iver & Farley, 2003)
5. Improving student achievement (Coburn & Talbert, 2005)
6. Communication (Chen, Heritage, & Lee, 2005)
7. Staff allocation (Honig & Coburn, 2008)

These responsibilities are defined in the following paragraphs.

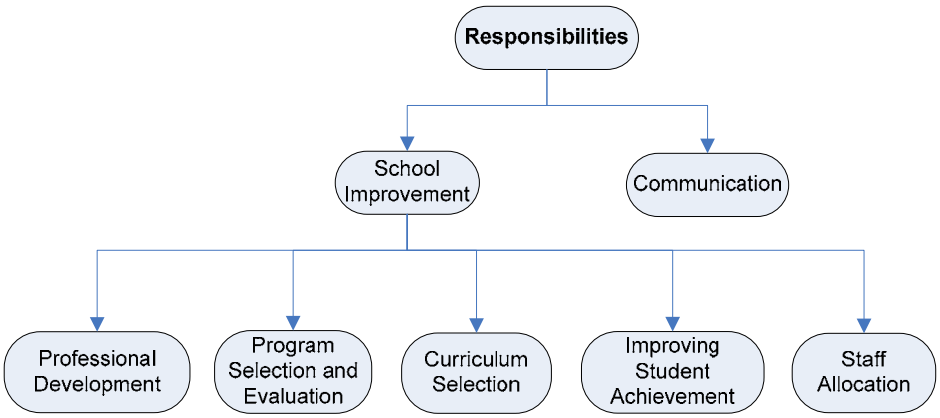

Figure 1. **Policymaker responsibilities.**

Administrators make use of the following information and evidence to enact *school improvement plans*: day-to-day information on student strengths and needs; outside information such as goals, strategies, community and political pressure, and partnerships to help them make decisions under ambiguous conditions; and surveys from parents (Honig, 2003) and students (Massell's study, as cited in Honig & Coburn, 2008). Sometimes improvement plans also serve as data for decisions about professional development, textbooks, and other district decisions as suggested by funding sources such as Title I (Honig & Coburn). Although administrators engage in confirmatory practices when searching for research to support approaches they take or intend to take, this practice can be used to support reforms that could contribute to school improvement (Honig & Coburn).

The broad term *professional development* is simply defined as a comprehensive, sustained, and intensive approach to improving teachers' and principals' effectiveness in raising student achievement (NSDC, 2008). Professional development activities for teachers and principals should be designed based on a clear set of learning goals and implemented by using evidenced-based learning strategies aimed at improving instructional effectiveness and student achievement. These activities may include learning about analyzing student performance data and using this information for development of formative assessments and other educational materials.

Program selection and evaluation helps administrators make decisions about how to direct funding and resources toward identified areas of need (Brunner et al., 2005), sometimes taking the form of legitimizing existing programs and decisions (Coburn & Talbert, 2005). In general, program evaluation focuses on the central idea of whether programs should be kept or replaced. Program activities need to be continually monitored for implementation and effectiveness and changed as needed (Honig & Coburn, 2008). This can be done by looking at program and student achievement data to see if progress is being made and, if not, what corrective action needs to be taken (Guerard, 2001).

Curriculum selection includes general decisions about curriculum adoption (Coburn et al., 2009; Coburn et al., in press; Honig & Coburn, 2008). Administrators also make decisions about curriculum beyond selection—for example, decisions about curriculum frameworks (Coburn et al., in press), the best curricular approach (Honig & Coburn), and linking curriculum and instruction to standards (Mac Iver & Farley, 2003).

Improving student achievement cannot be done without understanding what it is that students do and do not understand and distinguishing the students who understand from those who do not (Coburn & Talbert, 2005) . For example, performance data are used to place students into different performance categories, and then measures are taken to provide students with appropriate interventions. Evidence and activities that can be used to improve student achievement also include: examining student gains, making predictions based on data, identifying topics that students need help in, creating individualized education plans, and examining curricular decisions that have been made based upon trends such as student achievement (Coburn & Talbert).

Staff allocation. In order to improve students' performance, policymakers need to allocate and prepare the staff required to implement successful educational programs (Honig & Coburn, 2008).

Communication includes conversations as well as sharing information with all stakeholders. Communication occurs inside of the school building or school system among teachers, staff, students, and the district central office (Brunner et al., 2005; Honig & Coburn, 2008). One example of this type of communication is district central office and school staff dialogues for resource allocation (Honig & Coburn). Communication with individuals outside the school, such as parents and community members, can be in the form of reports (online or print) (Chen, Heritage, & Lee, 2005) or traditional parent-teacher conversations (Brunner et al., 2005).

Testing and Data Management Responsibilities

To bolster the literature on the responsibilities of policymakers and the types of decisions they make, two focus groups with district policymakers from various parts of the United States were held. Five policymakers participated in the first focus group: one from Maine; two from New Jersey (one superintendent and one curriculum and instruction supervisor), and two from New York (one director of information services and one retired assistant superintendent for curriculum and instruction). The second focus group had four participants: an administrator of support services from Arizona; an assistant superintendent from California; a curriculum coordinator from Oklahoma; and a coordinator of assessment from Texas.

We found that in addition to the roles and responsibilities that the research literature defined, a number of district-level roles relate specifically to testing. District people administer state, district, and (when it exists) benchmark testing, plus all the school-wide tests including the Advanced Placement Program® (AP®), ACT, SAT®, Preliminary SAT (PSAT), and all English as a second language (ESL) and special needs testing. They prepare trend data, subpopulation test performance at the school level, and multiple measure comparison reports between state and local results for principals so they can do their respective analyses. Their jobs include explaining how state accountability fits into district level accountability and training teachers and principals to access and interpret assessment data. Finally, it is their responsibility to select and integrate data management systems, including providing all the support necessary.

Data access is becoming more and more time-consuming for these stakeholders, who want to simply get relevant reports with the push of a button. The interviewees noted that some principals are intimidated by the data and that they need training. The infrastructure does not yet exist in most places to easily administer tests, scan answer sheets, and transmit reports to principals. In addition, it is expected that principals will share the data with teachers, students, and parents, but there are roadblocks for printing and delivering these reports in a timely manner. Some districts are actively working on improving this situation. Some are selecting and purchasing data management systems, and some are building "homegrown" solutions and data systems. They often lack staff members who are knowledgeable in these areas (including software engineering, interface design, report design, and statistics), so there is a lot of recreating the wheel, as well as training to get people up to speed, for district stakeholders as well as principals and teachers.

An updated depiction of district-level responsibilities is shown in Figure 2. This new information not only bolsters the information on responsibilities and decisions of the district-level policymaker, but it also highlights one of the main problems with using assessment data—namely, that there are difficulties in accessing and interpreting assessment data at all stakeholder levels. This provides support for designing an intuitive evidence-based reporting framework to facilitate decision making by enhancing access to available data (see Section 3). Next, we examine the roles that evidence plays in district-level policymaker decision making.

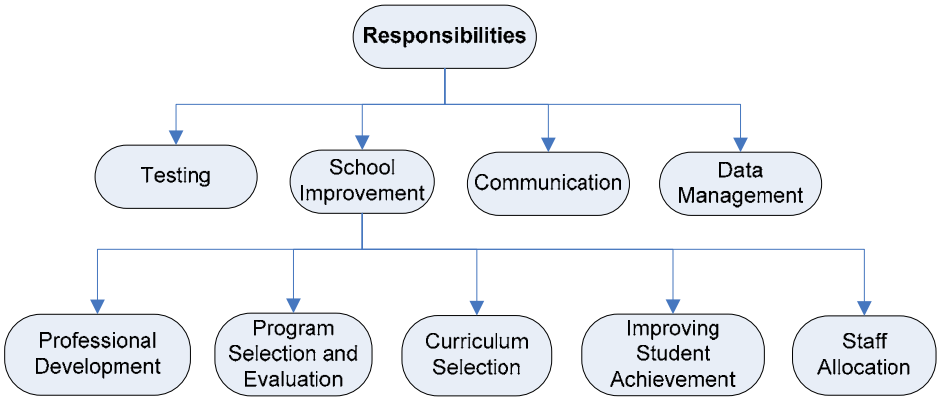

Figure 2. **Updated policymaker responsibilities.**

Use of Evidence

District offices have multiple constituencies to serve and multiple layers of governance to whom they must be responsive. Evidence, which can take many forms, has been identified as playing five roles in decision making: instrumental, conceptual, symbolic, sanctioning, and no role (Coburn et al., 2009). An *instrumental* role of evidence is one where administrators use evidence directly to provide guidance to decisions related to policy or practice. This rarely happens. In one in-depth analysis of 14 types of decisions made by district administrators in 16 districts (Coburn et al.), only 2 decisions appear to have been made by using data or evaluation research to directly inform decisions. Other studies cited by Coburn et al. report similar results. It is important to note that even when evidence does play an instrumental role, people sometimes interpret the results differently. In addition, and as would be expected, policymakers also consider budgetary, political, and administrative issues when making decisions.

A *conceptual* role of evidence is one that provides decision-makers with new ideas, concepts, or generalizations that influence how they view the nature of problems. As such, it sometimes provides background information rather than guiding particular decisions. This also rarely happens. Policymakers tend to search for and pay greater attention to evidence that resembles what they already know, while interpretation is influenced by individuals' pre-existing beliefs and experiences (Coburn et al., 2009).

A *symbolic* role of evidence is one that is used to justify pre-existing preferences or actions. The main function of this type of evidence is to create legitimacy for solutions that are already favored or even enacted. A typical pattern is examining literature selectively or recruiting experts who are advocates of the preferred strategy. In one study, evidence was used to justify decisions that were already made in 7 out of 14 decisions (Coburn et al., 2009, p. 15). Another study found symbolic uses of evidence in 4 out of 16 districts (Coburn et al., p. 15). In a third study, when there was a dip in test scores in the first year of instituting a new curriculum, which is predictable for a new curriculum, individuals who were opposed to the new curriculum used the dip to organize opposition to the curriculum, causing the district to stop using it.

A *sanctioning* role of evidence is one where evidence is used at one level (e.g., state or federal) to create a list of programs that are approved for use by units below them (e.g., district). Districts choose programs from this list in order to receive state or federal funding, but do not

review the evidence themselves. Given increased federal and state requirements that schools use "research-based" programs, we may see an increased role of this type of evidence.

Finally, districts often make decisions without reference to research, evaluation findings, or systematic data, which is being called *no role* of evidence. In an analysis of 35 decisions about Title I programs (cited in Coburn et al., 2009), 25% of the decisions were made on the basis of political or financial concerns alone, and another third of the decisions were based on impressions or anecdotal information. In another study, one out of three districts used evidence in choosing curriculum adoption and none used evidence while making decisions about professional development.

What Is Considered Evidence?

There are two types of evidence that policymakers use: evaluation studies of programs and student performance assessment data. We need to be aware of both types of evidence, though the focus in this review will be on the use of student performance assessment data.

Why do local policymakers not make objective use of evidence all the time? One answer is that district administrators often lack the right evidence that addresses the question or issue at hand, in a form they can access and use, at the time that they need it (Coburn et al., 2009). Even when they have the relevant data, the data are not always in a form that allows district administrators to answer the questions that they have, or are simply too complex (Hambleton & Slater, 1994, 1996).

In addition to problems in the use of evidence, policymakers also have trouble accessing and making sense of assessment data in the form they currently receive it. For example, policymakers have difficulty reading and interpreting the reports they receive (Hambleton & Slater, 1996). Administrators misinterpret the meaning of symbols and specific terms used in assessment reports (e.g., statistical language), and the complexity of reports often causes additional confusion. Compounding this problem is the lack of time policymakers have to read and interpret assessment reports. As a result, the reports that policymakers get do not objectively inform their decisions. We will now describe a reporting framework that can be used to improve access to and use of student achievement data by policymakers.

Evidence-Based Reporting Framework

While the previous section mentioned two types of evidence, our framework will focus only on student achievement data and how score reports can be designed around these data to support policymakers' decision making. This reporting framework was inspired by work on evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003). ECD is a methodology for assessment design that emphasizes a logical and explicit representation of an evidence-based chain of reasoning from tasks to skills, with the goal of ensuring the validity of assessment results. Our framework links student achievement data to questions, which serve as a proxy for the decisions that policymakers need to make. Additional work on extending ECD principles to program evaluation has been presented elsewhere (Shute & Zapata-Rivera, 2007).

Our approach begins by creating a mapping from policymakers' questions (in support of decisions) to the student achievement data needed to answer such questions (see Figure 3). Policymakers may use a question-based interface to access student achievement data. A user profile keeps track of the questions the policymaker recently used, and also allows identification of preferred questions. Student achievement data may include information about general latent variables of interest (e.g., assessment claims regarding student competency in content areas aggregated across schools), as well as information concerning the reliability and validity of inferences drawn from the assessment data. Reports addressing a particular question are produced in a form that is geared toward the stakeholder. These results can either directly inform decisions to be made in the future (instrumental role of evidence) or can spark new questions to give other perspectives about the data (conceptual role of evidence).

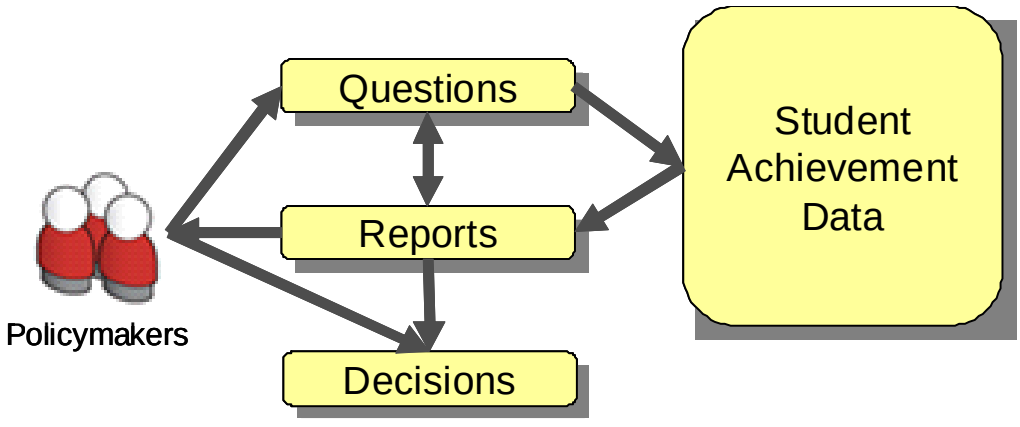

Figure 3. **Evidence-based reporting framework.**

As we have seen, policymakers have important responsibilities related to school performance. Student achievement data can be used in many ways to inform the decisions related to these responsibilities. However, policymakers currently receive achievement reports that are not designed to inform those decisions; even when they are, they are difficult and time-consuming to read and interpret. When they do use assessment data, they generally "mine" the results, which is more in line with the symbolic uses of evidence described earlier in order to justify predetermined decisions. To use evidence objectively and well, it helps to begin with a question you want answered and to link those questions from the assessment(s) and other data gathered back to the decisions. The next section attempts to enumerate the types of questions policymakers ask and how this reporting framework can be used to support evidence-based decision making.

Our approach makes connections from questions to decisions via actions and evidence using ECD principles. We will now describe results from a literature review and interviews with policymakers about the questions they ask to support their decision making.

Policymaker Questions

What questions do policymakers ask? To answer this, we reviewed the literature in areas such as data-driven decision making, achievement data for policymakers, and educational policy. We also reviewed a special issue of the *Journal of Education for Students Placed at Risk*, Volume 10, Number 3, which has a number of articles on the subject.

Relevant articles from other journals were selected and used to characterize the types of decisions policymakers make as part of their responsibilities, as well as the questions they tend to ask in support of these decisions. These articles include: Brunner et al. (2005); Coburn et al. (2009); Coburn et al. (in press); Englert et al. (2004); Guerard (2001); Hambleton and Slater (1996); Honig and Coburn (2008); Mac Iver and Farley (2003); Snipes et al. (2002); and Streifer and Schumann (2005).

In addition, several district policymakers from various parts of the United States were interviewed, as was described earlier, about the responsibilities they have and the questions they would like answered by assessment data to help them make decisions.

We identified two overarching types of questions policymakers ask: those related to knowing about student achievement and those more directly related to making decisions based

on available information. Each of these has a number of categories that form their own respective hierarchies (see Figure 4). Student achievement questions are typically inquiries into performance data. There are three components of student achievement questions: the group of students (who), the types of data desired (what), and the content areas of concern (competencies). There are six groups of students, ranging from the entire state population of students to an individual student. The types of data can be either scores or performance levels, and these can be viewed either as distributions, comparisons over time of one group, comparisons to other groups (e.g., the state or other districts), or in terms of strengths and weaknesses. Sample questions are shown for each student achievement category (see Table 1).

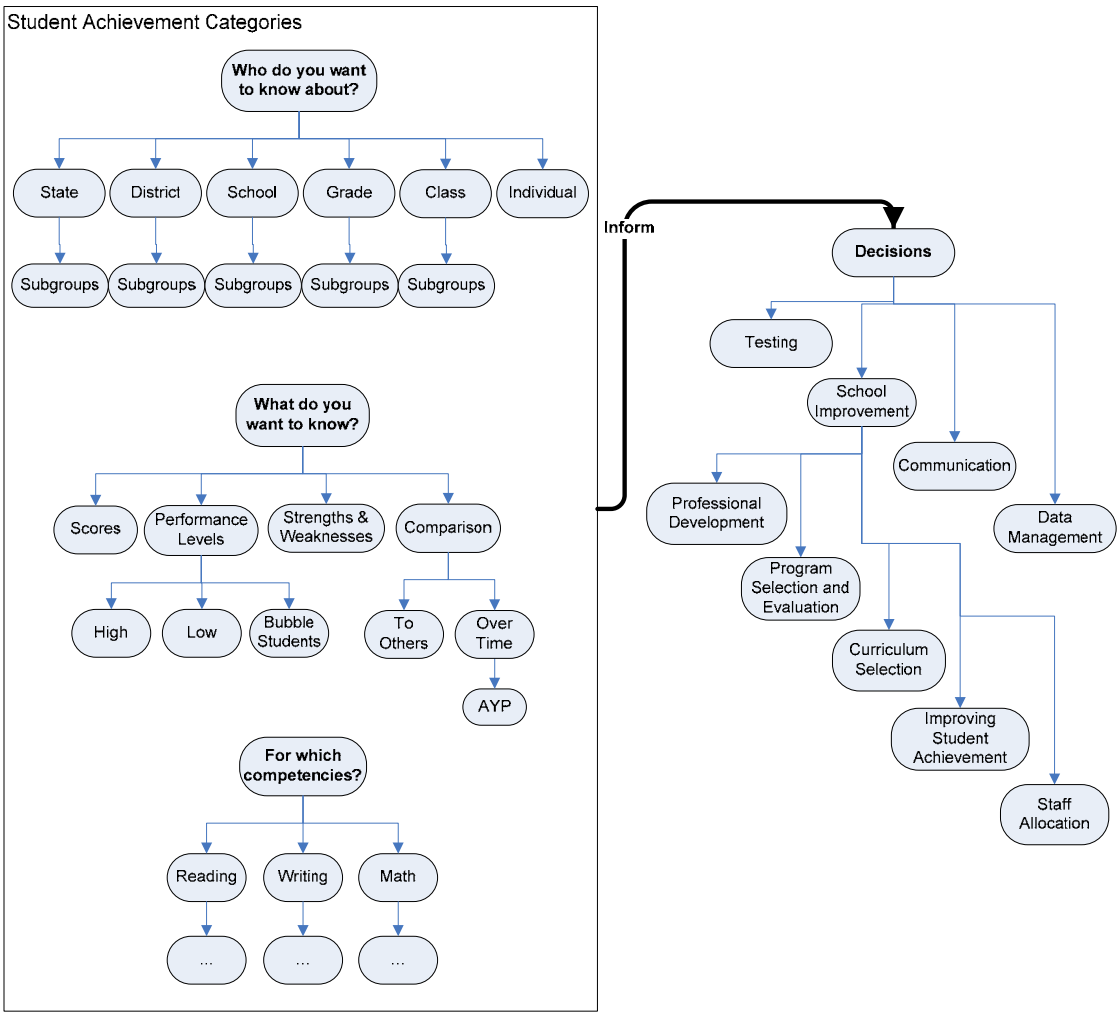

Figure 4. **Hierarchies of policymaker question types.**

Note. AYP = adequate yearly progress as defined by the No Child Left Behind (NCLB) act.

Table 1

Student Achievement Data Categories and Sample Questions

	Who	What	Competencies	Sample question
1	District	Performance level	All	How is my district performing on the competencies?
2	District	Performance level	Reading	How are my district's students performing in reading?
3	District subgroups	Performance level	All	Are there any narrowing differences in academic achievement between white and minority students?
4	District	Comparison, to others	All	How does our district compare to other districts in the state?
5	District	Comparison, over time	All	Has the district shown any improvement over time?
6	District, by school	AYP	All	Which schools need help to meet AYP goals?
7	District, by school	Strengths and weaknesses	All	Are my schools weaker in some areas?
8	District	Performance level, bubble students	All	How many students were 1-2 questions from acceptable performance?

Note. AYP = adequate yearly progress as defined by the No Child Left Behind (NCLB) act.

The decision questions are typically related to policymakers' responsibilities and can (and probably should) draw on student achievement data in order to make objective decisions. These categories are defined in an earlier section. Sample questions are shown for each category (see first two columns of Table 2). Since the goal is to find student achievement data to help with decisions, we do not include examples for data management (decisions in this area are about professional development of administrators and others who are not directly connected to student achievement).

Table 2

Decision, Questions, Recommended Data, and Suggested Actions

	Decision category	Questions	Recommended data	Suggested actions
D1	Professional development	What professional development (PD) can I offer to help them relative to achieving state standards?	District level student performance levels and strengths and weaknesses on competencies Drill down to schools Drill down to subgroups	Identify PD relative to performance levels, identify high and low performing groups
D1	Professional development: best practices	Are there any schools and classes that are doing so well that they can serve as a model of best practices for others to replicate?	Schools and classes—high performance	Identify best practices from consistently high performing teachers to guide professional development
D2	Staff allocation	How should staff be allocated to improve student achievement?	Schools—strengths and weaknesses	Match teachers with students based on teachers' areas of expertise and students' needs
D3	School improvement	How can I improve low test score in math in middle school?	School—low performance (subject area)	Select a new intervention, instructional approach, or program
D4	Program selection and evaluation	Are specific programs/practices improving student achievement?	Change in academic achievement since program instituted	Evaluate programs in terms of student growth Determine future program selection and retention, and opportunities for professional development
D4	Professional development	What are the instructional strengths and weaknesses of schools and what instruction should be changed accordingly?	School strengths and weaknesses	Identify consistently low-performing teaching areas to guide professional development for better teaching practices
D5	Curriculum selection	What types of changes should I be making in curriculum to see improvements?	Student achievement areas of weakness and strength Growth since curriculum instituted	Choose curriculum as informed by continued student areas of weakness and strength

	Decision category	Questions	Recommended data	Suggested actions
D5	Curriculum selection	Are there any instructional resources available that are aligned to the instructional priorities?	Student achievement areas of weakness and strength	Choose appropriate instructional resources as informed by student areas of weakness and strength
D6	Communication	How do I present student achievement data to stakeholders?	District—strengths and weaknesses	Interpretation and dissemination plans customized to each type of audience
D6	Communication	How to help teacher to get to know students better?	Individuals—strengths and weaknesses	Individualized learning plans, individual conferences
D7	Test evaluation	Are the results an accurate reflection of the achievement of the school overall?	District—student achievement	Test evaluation, compare achievement data to other evidence, note surprising results
D7	Test evaluation, school improvement	Have the students had the opportunity to learn the curriculum or standards assessed?	District—strengths and weaknesses	Identify gaps in instructional sequence as compared to test coverage Compare low achieving areas and the curriculum and instructional sequence Be aware of student mobility

Note that student achievement questions do not always suggest decisions and, likewise, questions asked in support of decisions do not always suggest readily available student achievement data. We assert that there are implicit links between decision questions and student achievement data; we supply these links as suggestions in our reporting system to help policymakers make objective decisions. This effort is described in the next section.

Linking Questions to Assessment Data and Decisions

While policymakers make no distinction between questions about student achievement (SA) and decisions (D), in that they are all questions they would like to have answered, we hypothesize that answers to SA questions can inform D questions (as represented by the *Inform* arrow in Figure 4). In many cases, however, when SA questions are asked, policymakers may

have no D's in mind. Instead, they look for patterns in the data to support existing decisions about school improvement activities. This approach cannot always lead to fruitful (i.e., improvement in student performance) decisions. In order for administrators to make objective decisions about how student performance can be improved, they need to link student achievement data more directly to decision efforts.

For each of the sample decision questions, we identified student achievement data that could help provide answers. This was an iterative process where three researchers individually identified useful data for a particular decision, and then reached consensus about the most pertinent ones. Using the same process, we also identified next steps (i.e., suggested actions) that could help with the decisions (see Table 2). There was an interesting pattern in the questions, namely, that for any school improvement plan (e.g., program selection, professional development), two types of causal questions are asked: the first about how to improve student achievement (i.e., predictive questions) and the other about how to assess programs that are already in place (i.e., evaluative questions). This distinction will help us make decisions about the types of actions that can be suggested.

For each sample student achievement question, we refer to the decision-question analysis to suggest the types of decisions that can be made from those data. The reports that appear in response to a user selecting an SA question will leverage these links. In particular, each report will show a graphical depiction of the data that can answer the selected question. To supplement the understanding of each graphical depiction, there will be a textual description of the results, highlighting of main results and anomalies, as well as the limits of what can be interpreted (this will include statistical information). This alone is an improvement over existing reports, where the majority of the interpretation is done by the user. To truly provide additional value for these reports, there will be suggestions for how the data can be used for decision making, as well as suggestions for additional data (via questions) to help refine the evidence for decision making.

In the next section, we show an example scenario using a prototype reporting system for student achievement data questions such as those in Table 1, along with suggested interpretations and possible actions the policymaker might take based on those data.

Example: Exploring Assessment Data

The following example is based on the CBAL project (Bennett & Gitomer, 2009). CBAL's theoretically driven assessment design makes use of cognitive-scientific principles, competency models, and developmental models. CBAL uses Periodic Accountability Assessments (PAAs) that can provide more information in terms of both depth and breadth of student knowledge in relation to curriculum standards. Each PAA acts as a piece of a hypothetical long test. Several PAAs are administered across the school year. Feedback for teachers and students is provided throughout the year. A final accountability result is derived by aggregating performance from each PAA.

At the end of a hypothetical year, the middle schools in District X have just completed their final CBAL PAA in mathematics, reading, and writing. District Superintendent Brown wants to view the data to see how the schools in the district are performing. First, he wants an overall picture of their performance.

Mr. Brown can choose to access data according to the student achievement categories or by the decision categories (see Figure 4) and, in either case, can choose a question that will bring up an appropriate report. In our example, he selects a question from the student achievement categories (e.g., "How did the schools in my district perform on the eighth grade PAA?"). Each question is linked to a predefined query that makes use of assessment data available in the evidence-based reporting framework. Mr. Brown's user profile gets updated with the question(s) that he has explored. The report in Figure 5 shows a graphical depiction of the data[1] he requested (e.g., sorted by school, overall district scores), the data type (e.g., percent of students at each performance level), and the PAA competencies (e.g., reading, writing, and mathematics). This representation shows the percentage of students performing at the proficient level or above, as well as below. The report also includes a textual summary of the results, highlighting results and anomalies, along with acceptable interpretations of the data with a focus on appropriate decisions.

This report also suggests questions to explore that can provide evidence for particular actions and decisions the policymaker might make.[2] Clicking on links in the *Look at* field (see bottom of Figure 5) will take Mr. Brown directly to another report. The bracketed fields in the *Look at* field indicate variables in the query, leveraging the three components of student achievement questions identified earlier: the group of students, the types of data desired, and the

content areas of concern. Additional variables will be added as needed (e.g., the number of years for comparisons over time).

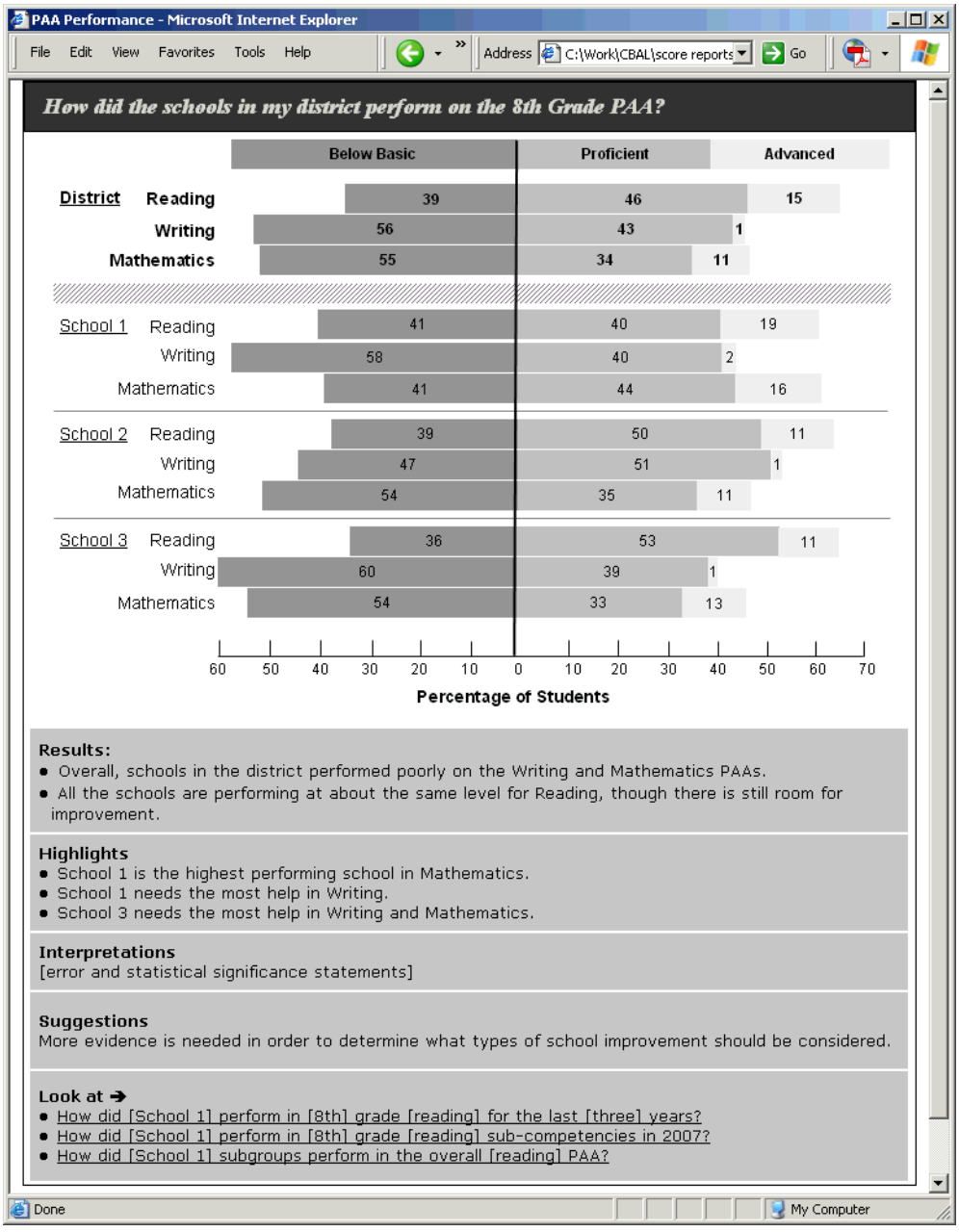

Figure 5. **Sliding bar plot of District X final eighth grade PAA performance data for reading, writing, and mathematics.**

This report shows relatively poor performance in all three PAA competency areas across the district, as well as in the individual schools. Since reading is a skill needed for the other competencies, and as the report points out, there is room for improvement in reading even though it is not the area of the lowest performance, Mr. Brown decides to look at the trend data for reading for School 1 grade 8 (e.g., over the last 3 years). The report in Figure 6 appears in response.

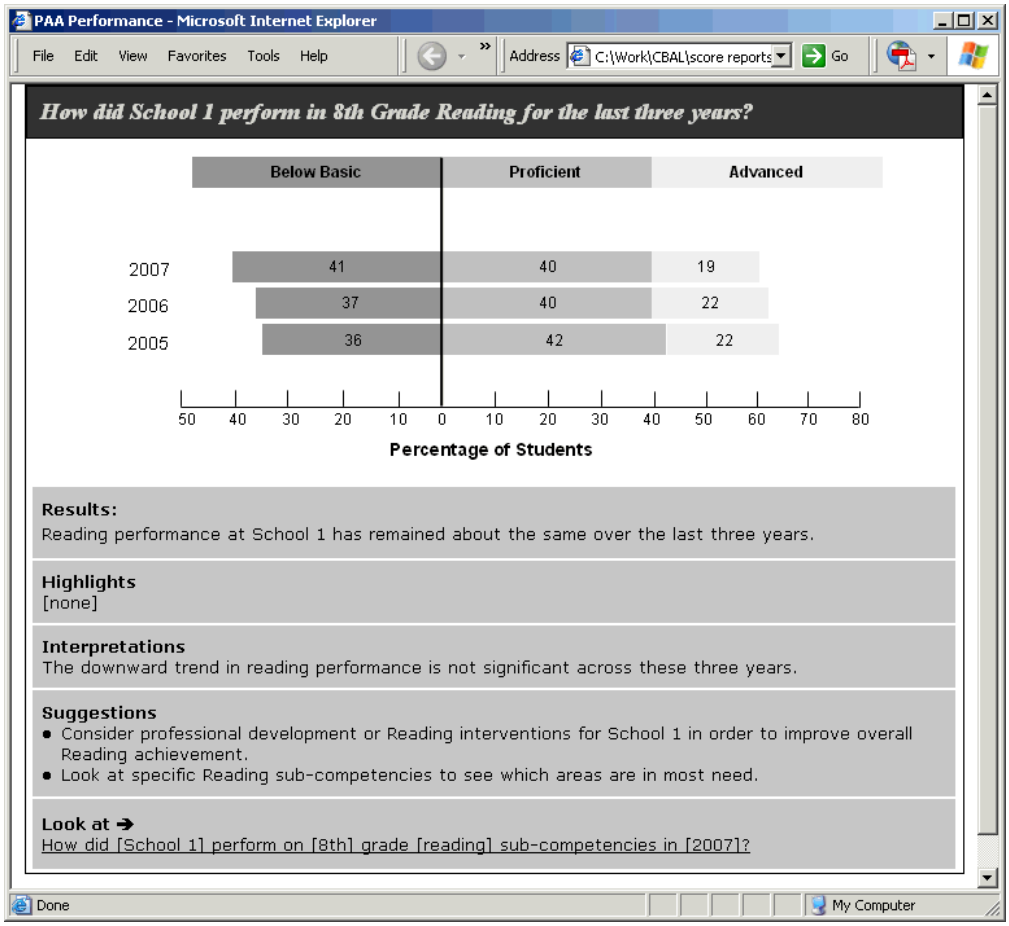

Figure 6. **Sliding bar plot of School 1 Grade 8 reading performance data, 3 years.**

This second report shows the same sliding bar representation, carefully using the same colors and layout as the previous report so that they are easy to read and compare. In this case, Mr. Brown sees that their performance seems to be getting worse in reading for School 1 over the previous 3 years, but the *Results* field points out that the reading scores have remained about the same, and the *Interpretations* field notes that the perceived change is not significant. Decisions should not be made based on these changes in scores.

Mr. Brown decides to look at School 1's reading subcompetency performance (see Figure 7). The same representation is still used, this time reporting performance at only two levels: *Needs Help* and *On Track*. The *On Track* level combines the previous proficient and advanced levels, and uses a color somewhere between the other two (in this case, an in-between gray) for distinction. Mr. Brown sees that the vast majority (92%) of students are on track with their basic reading skills, which is a good thing.

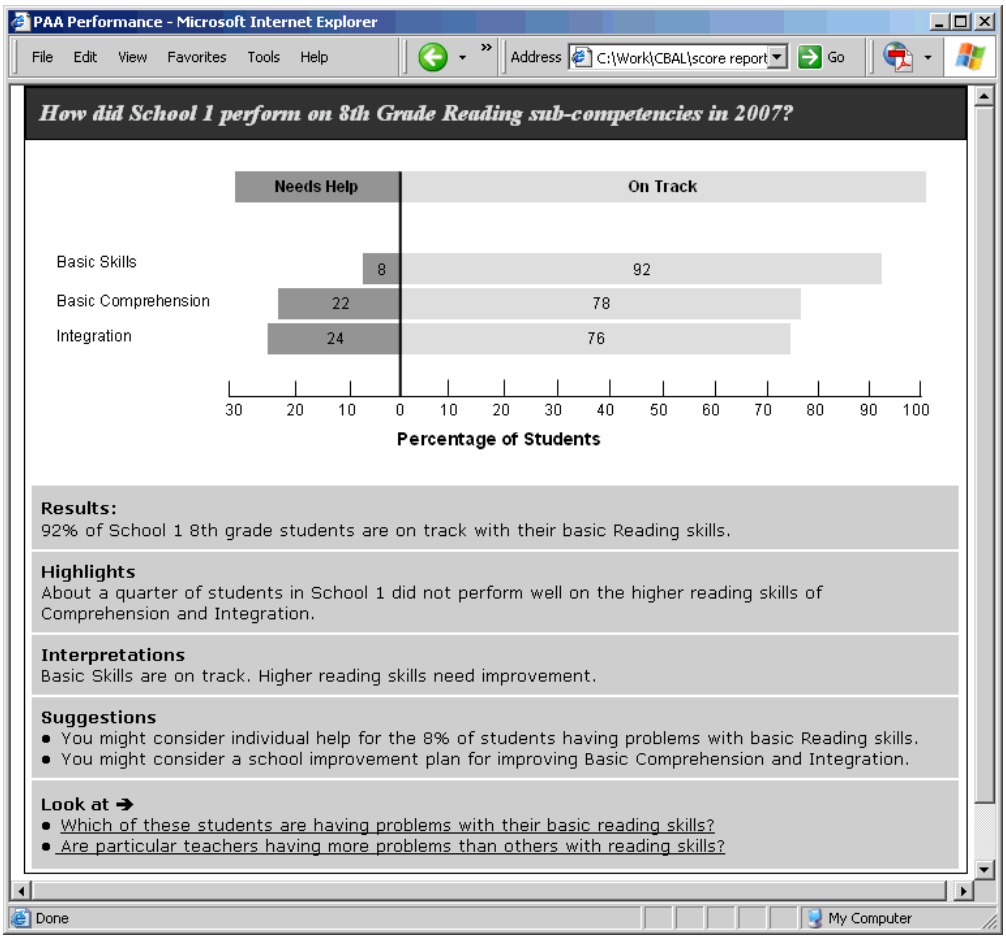

Figure 7. **Sliding bar plot of School 1 Grade 8 final detailed reading performance data.**

Mr. Brown decides to explore how subgroups performed on the reading PAA. He goes back to the previous page (Figure 7) and clicks on the appropriate *Look at* question. The results for this question are presented in Figure 8.

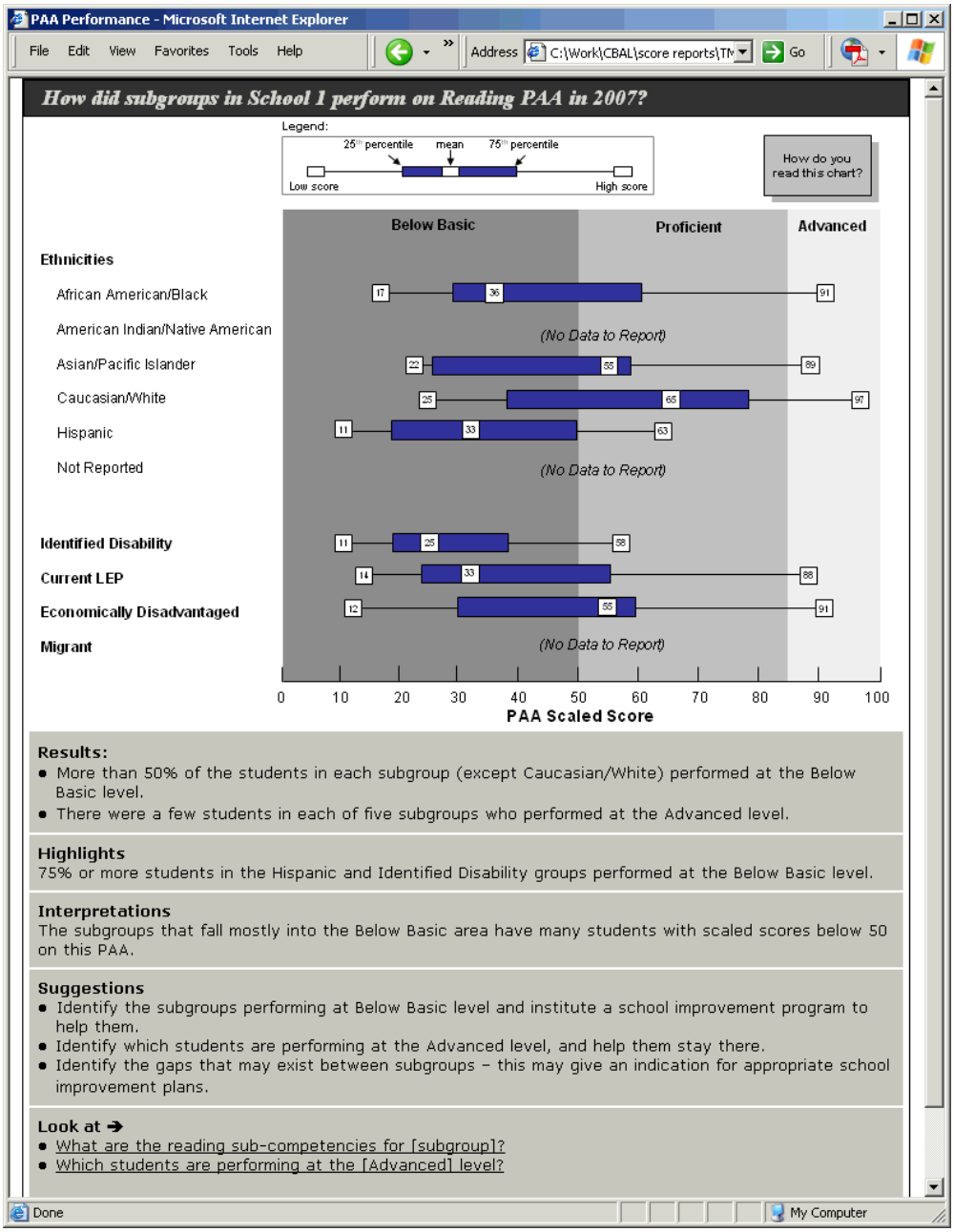

Figure 8. **Box-and-whiskers plot of subgroup performance in School 1 Grade 8 reading.**

The representation of this fourth report is a box-and-whiskers plot, a succinct way to report the distribution of scores and useful when there are many groups to compare. This chart shows the end-of-year performance of the state's NCLB subgroups for eighth grade students in School 1 for reading. The blue (or dark gray, as perceived in gray scale) rectangle part of each line shows the middle 50% of the student scores; the lines on each side show the bottom and top

25% of student scores. The median score is shown in the rectangle, and the lowest and highest scores appear to the left and right of the line, respectively. Note that the legend appears above the chart, since focus groups have shown that few people look for them when they appear underneath the chart. The “How do you read this chart?” button will pop up in a window when clicked to walk the user through an example. This representation may be difficult to understand and interpret at first glance by the uninitiated, but research has shown that with a little instruction policymakers prefer this concise representation (Hambleton & Slater, 1994).

Mr. Brown realizes that there needs to be more emphasis on helping some of these subgroups improve in school. He has been leaning toward professional development for other reasons, and this is giving him additional evidence for this decision, so he makes a note to talk to the principal and share these results with her. He will continue to look at the student achievement data and pass results on to the appropriate people. Mr. Brown finds it easy to access to the data using questions and appreciates the additional information provided to help him interpret the data.

User profiles maintain a list of questions for each user or user type. These questions can be used as starting points to look for particular answers and can be customized for each user based on his or her preferences.

Discussion and Summary

Policymakers need to get access to the right data in ways that facilitate decision making. This evidence-based reporting framework is a good first step toward supporting the instrumental and conceptual roles that evidence plays in making decisions. However, given the diversity of data sources that are used and the social complexity of the decisions that need to be made, it is likely that policymakers will still continue using evidence for various functions even when or if a reporting system such the one described here was available.

Even though other sources of evidence (e.g., research reports and results from program evaluation studies), politics, and other external forces are involved in decision making, we believe this reporting framework could help policymakers make objective decisions by using valid assessment data.

The use of a question-based interface that connects assessment information to the needs of policymakers may be a reasonable approach to interacting with assessment information that

could potentially be applied in other contexts to support decision making. It is the aim of this framework to support sound, transparent, evidence-based decision making.

This paper presented an evidence-based reporting framework for helping district-level policymakers find and use assessment information that is appropriate for their needs. An example based on the CBAL project (Bennett & Gitomer, 2009) was used to illustrate this approach. Future work includes designing a variety of score report prototypes and refining the reports based on feedback gathered from district-level stakeholders.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional development. In L. Darling-Hammond & G. Skyes (Eds.), *Teaching as the learning professional: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. Cumming (Eds.), *Assessment issues of the 21st century* (pp. 43-61). New York, NY: Springer.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mardinach, E., & Fasca, C. (2005). Linking data and learning: The Grow Network study. *Journal of Education for Students Placed at Risk, 10*(3), 241-267.

Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students' learning needs with Technology. *Journal of Education for Students Placed at Risk, 10*(3), 309-332.

Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What's the evidence on districts' use of evidence? In J. Bransford, D. J. Stipek, N. J. Vye, L. Gomez, & D. Lam (Eds.), *The role of research in educational improvement* (pp. 67-88). Cambridge, MA: Harvard Education Press.

Coburn, C. E., & Talbert, J. E. (2005, April). *Conceptions of evidence use in school districts: Mapping the terrain.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Coburn, C. E., Toure, J., & Yamashita, M. (in press). Evidence, interpretation, and persuasion: Instructional decision making in the district central office. *Teachers College Record, 111*(4).

Englert, K., Fries, D., Goodwin, B., Martin-Glenn, M., & Michael, S. (2004). *Understanding how principals use data in a new environment of accountability.* Aurora, CO: Mid-continent Research for Education and Learning.

Guerard, E. B. (2001, May 17). School leaders learn to make data-driven decisions. *eSchool News*. Retrieved from www.eschoolnews.com/news/showstory. cfm?ArticleID=2621

Hambleton, R. K. (2007, June). *A new challenge: Making test score reports more understandable and useful.* Paper presented at the CCSSO annual meeting, Nashville, TN.

Hambleton, R. K., & Slater, S. (1994). Using performance standards to report national and state assessment data: Are the reports understandable and how can they be improved*?* In L. Crocker & M. Zieky (Eds.), *Proceedings of Joint Conference on Standard Setting for Large-Scale Assessments* (Vol. II, pp. 325-343). Washington, DC: U.S. Government Printing Office.

Hambleton, R. K., & Slater, S. (1996, April). *Are NAEP executive summary reports understandable to policymakers and educators?* Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

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), 292-338.

Honig, M. I., & Coburn, C. E. (2008). Evidence-based decision-making in school district central offices: Toward a policy and research agenda. *Educational Policy*, *22,* 578-608.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, *33*(8), 16-30.

Mac Iver, M. A., & Farley, E. (2003). *Bringing the district back in: The role of the central office in improving instruction and student achievement* (CRESPAR Report No. 65). Baltimore, MD: Johns Hopkins University.

Miller, N. (2003, February). Best practices. *Super Tech News*, *2*(1). Retrieved from http://www.blegroup.com/supertechnews/feb03.htm#best

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, *1*(1), 3-62.

National Staff Development Council. (2008). *NSDC's definition of professional development*. Retrieved from http://www.nsdc.org/standfor/definition.cfm

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4-14.

Shute, V., & Zapata-Rivera, D. (2007). *Evidence-centered evaluation (ECE): Capturing, weighing, and understanding research to improve achievement for Hispanic students.* Unpublished manuscript.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement.* Washington DC: The Council of the Great City Schools. Retrieved from http://agi.harvard.edu/events/download. php?id=90

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

Spillane, J. P. (2004). *Standards deviation: How local schools misunderstand policy.* Cambridge, MA: Harvard University Press.

Spillane, J. P. (2005). Primary school leadership practice: How the subject matters. *School Leadership & Management, 25*(4), 383-97.

Streifer, P. A., & Schumann, J. A. (2005). Using data mining to identify actionable information: Breaking new ground in data-driven decision making. *Journal of Education for Students Placed at Risk*; *10*(3), 281-293.

Wayman, J. C., Midgley, S., & Stringfield, S. (2005, April). *Collaborative teams to support data-based decision making and instructional improvement.* Paper presented at the 2005 annual meeting of the American Educational Research Association, Montreal, Canada.

Notes

[1] These data are all fictitious.

[2] In this report, we are focusing on appropriate representations for answering the selected questions to help policymakers make decisions.