How Did That Happen? Teachers’ Explanations for Low Test Scores

MARGARET EVANS
Illinois Wesleyan University

REBECCA M. TEASDALE
University of Illinois at Urbana-Champaign

NORA GANNON-SLATER
Breakthrough Charter Schools

PRIYA G. LA LONDE
Georgetown University

HOPE L. CRENSHAW
Teen Health Mississippi/Mississippi First

JENNIFER C. GREENE
University of Illinois at Urbana-Champaign

THOMAS A. SCHWANDT
University of Illinois at Urbana-Champaign

Context: Educators often engage with student performance data to make important instructional decisions, yet limited research has analyzed how educators make sense of student performance data. In addition, scholars suggest that teachers recognize a relationship between their instruction and student performance data, but this is a relatively untested assumption.

Focus of Study: We investigated if and how teachers referenced instruction as a contributing factor for why students performed in particular ways on assessments. We also studied other explanations that teachers offered for student performance data.
Research Design: Our research team conducted a qualitative case study of six grade-level teams of teachers who met biweekly to make meaning of student performance data. Using data collected from 44 hours of observation of teacher team meetings, 16 individual interviews, and six group interviews with participating teachers, we analyzed the ways in which and the extent to which teachers referenced instruction as a contributing factor to student performance data. Findings: Teachers connected student performance data to their instruction approximately 15% of the time. Teachers more frequently connected student performance data to student characteristics. Notably, student behavior accounted for 32% of all teacher explanations for student performance. We offer five distinct categories of teachers’ explanations of student performance and the extent to which teachers invoked each category.

Conclusions: The findings in this study build on research on teachers’ attributions for assessment data. In contrast to other studies, our findings suggest that teachers invoked student characteristics in distinct ways when explaining student performance. At times, teachers were knowledgeable about student characteristics, which offered verifiable insights into the “problem” of low achievement. At other times, teachers voiced negative viewpoints of students that served to blame students for their poor performance. We suggest that the practice of data-driven decision making offers an opportunity to bolster educators’ informed judgment and undermine negative, unverifiable claims about children.

Given the substantial interest in data-driven decision making (DDDM) in K–12 schools, many teachers now have an explicit requirement to consider assessment data on students’ performance when making instructional decisions (Bertrand & Marsh, 2015; Dunn, Airola, Lo, & Garrison, 2013; Marsh, Pane, & Hamilton, 2006). Student performance data are often described as information that teachers can use to “drive instruction” (Duncan, 2009). Yet, the jury is out on whether instruction can be shaped in this way.

Some researchers and educational leaders suggest a tight-knit relationship, where data offer clear and objective answers to questions of what and how teachers should teach. For example, former Secretary of Education Arne Duncan (2009) argued that:

Our best teachers are using real-time data in ways that would have been unimaginable five years ago. They need to know how well their students are performing. They want to know exactly what they need to do to teach and how to teach. . . . They aren’t guessing or talking in generalities anymore.

This sentiment that data take the “guessing” out of instructional decision making is widely advanced in DDDM research and policies (Duncan, 2009; Mandinach, 2012; Means, Padilla, & Gallagher, 2010). However, several researchers are critical of the view that assessment data alone can inform instructional decisions. They claim that assessment data are but one of multiple sources of information that educators may draw from
to support their instructional decisions (Coburn & Turner, 2011; Dowd, 2005; Johnson & La Salle, 2010). So, although the extent to which data directly inform educators’ decisions is debated, scholars generally agree that student performance data can be informative, even important, for instructional decision making (Coburn & Turner, 2012; Kvernbekk, 2011; Mandinach, 2012).

Although there is consensus that teachers ought to both examine and then act on this relationship between student performance data and their instruction, the critical questions are whether and how they actually do so in their daily practice. The existing research often fails to capture how teachers engage with data in their practice because little research includes direct observation of teachers’ engagement with data in their workplace (Coburn & Turner, 2012; Little, 2012). As Coburn and Turner (2012) stated, “We still have shockingly little research on what happens when individuals interact with data in the workplace setting” (p. 99). To address this gap in the research, our team observed practicing teachers during their workday at the times and spaces where they normally made sense of student performance data. The broad aim of this study was to investigate, when practicing teachers examine student performance data, what do they discuss? What decisions do they make? And finally, how does teachers’ knowledge of students come to bear in teachers’ data use conversations?

Our research team investigated these questions by observing a sample of practicing teachers during their workday as they discussed student performance data. Over the course of an academic school year, we observed six teams of teachers in Grades 3–5 as they analyzed and deliberated the meanings of a variety of types of assessment data on students’ performance. Toward the end of the study, we conducted individual and group interviews with participating teachers to clarify our understandings of what occurred during their data talk discussions. The particular analysis reported herein addressed the following research question: In what ways, if any, do teachers connect student performance data to their instruction?

In theoretical portrayals of teacher data use, educators examine data to make changes or improvements to their practice, which assumes that teachers perceive a link between students’ performance and their work as educators (Coburn & Turner, 2011; Marsh, 2012). Also, in educational policy discussions on teacher data use, practicing teachers are expected to directly link student performance data to instruction (Duncan, 2009; Mandinach, 2012). Yet, in practice, data are lacking on the connections that teachers actually do or do not make between their instruction and student performance data.

We aim to contribute to a growing body of research that investigates if and how teachers make connections between students’ academic
performance and their own instruction, daily practices, and decisions. To home in on the relationship between students’ performance data and teachers’ instruction, we studied the explanations teachers offered for why they thought students performed in particular ways. In the next section, we describe our intentional choice of studying teachers’ “explanations,” as opposed to teachers’ attributions.

EMPHASIS ON EXPLANATIONS OF STUDENTS’ PERFORMANCE

The literature on DDDM suggests two relationships between teachers’ instruction and students’ academic performance. In one view, students’ performance data inform instruction or offer teachers insights on “how to teach” (Duncan, 2009). Essentially, teachers can use student data to inform their future instructional decisions, a concept referred to as data use for teaching (Mandinach & Gummer, 2016, p. 2). For example, teachers may examine assessment data and locate a group of students who did poorly identifying the main idea of a reading passage. Teachers could then plan a new or revised lesson that aims to reteach those students how to identify the main idea of a passage. In this way, student data inform or “drive” teachers’ instruction.

Second, in contrast to data driving instruction, teachers’ instruction is also thought to “drive” or produce data. In other words, students’ performance data are viewed as representing the effectiveness of teachers’ past instruction (Duncan, 2009; Mandinach, 2012). For example, students’ performance on a reading comprehension assessment could offer insights into how effectively teachers previously taught reading comprehension skills.

The popular discourse that undergirds DDDM assumes that teachers perceive a relationship between their past instruction and students’ performance (Mandinach, 2012; Means et al., 2010), but this is a relatively untested assumption. In the current study, we specifically investigated whether this type of relationship exists in practice, that is, whether teachers consider their instruction as a factor contributing to students’ performance on assessments. We were also interested in what other explanations teachers offered for the performance data they reviewed.

The research on teachers’ attributions suggests that teachers do not consistently relate student performance to their instruction. For example, multiple studies have shown that teachers invoke students’ level of effort as a common explanation for why students perform poorly (Cooper & Burger, 1980; Jager & Denessen, 2015; Wissink & de Haan, 2013). Cooper and Burger (1980) found that teachers identified students’ “typical effort” as more influential on student performance than instruction (p. 98).
Similarly, Jager and Denessen (2015) found that teachers more frequently attributed low student performance to characteristics like students’ level of attention and effort than to teachers’ quality of instruction (p. 524). Other studies suggest that teachers may attribute low student performance to students’ race, gender, and/or ability level (Clark & Artiles, 2000; Reyna, 2000; Wissink & de Haan, 2013). For example, Reyna (2000) cautioned that race-based stereotypes can become the explanation for why students of color perform poorly in schools. Wissink and de Haan supported this caution with their findings, which suggest that teachers more readily attribute student performance to ethnic minority students’ lack of effort or ability as compared with students who belong to the ethnic majority. Overall, contrary to theoretical portrayals of DDDM, the research on teachers’ attribution challenges the assumption that teachers consistently relate student performance to their instruction.

Prior research strongly suggests that teachers’ perceptions of the relationship between their practice and student performance are critical determinants of how teachers proceed instructionally (Barrett, 2009; Booher-Jennings, 2005; Gallimore, Ermeling, Saunders, & Goldenberg, 2009; Gandha, 2014; Oláh, Lawrence, & Riggan, 2010). When teachers do not see their work as a contributing factor to students’ performance, then they may not consider student data as informative for their instruction. For example, Barrett (2009) and Booher-Jennings (2005) identified instances when teachers attributed a student’s low performance to the student’s inability or unwillingness to learn. These scholars argued that this particular explanation influenced teachers’ decisions to not offer these struggling students additional help, because they did not think students would or could take advantage of it. In contrast, Gallimore et al. (2010) gave an account of teachers who did connect students’ low performance to particular teaching strategies, and, as a result, these teachers were likely to try new teaching strategies for these students.

Although multiple scholars have suggested that teachers’ conceptualization of the relationship between students’ performance and teachers’ instruction is important (Coburn & Turner, 2011; Mandinach, 2012; Means, Chen, DeBarger, & Padilla, 2011), few studies on DDDM have investigated it directly. The present study aimed to build on the work of Oláh et al. (2010), and Bertrand and Marsh (2015), who used teacher interviews to empirically examine if teachers related student performance to their instruction. These scholars identified categories for the ways in which teachers sought to understand why students performed in particular ways. For example, Bertrand and Marsh (2015) found that teachers attributed students’ performance to four main factors: student characteristics, student understanding, the nature of the test, and instruction. Oláh et al. (2010)
found that teachers made a “diagnosis” for student performance and identified three categories of diagnoses related to instruction: students’ cognitive weaknesses, students’ procedural understanding, and students’ conceptual understanding. These scholars also identified one type of diagnosis unrelated to instruction, which they called contextual/external factors. Like Bertrand and Marsh (2015) and Oláh et al. (2010), our research also identified categories for teachers’ interpretations of the causes of student performance.

In contrast to the studies conducted by Oláh et al. (2010) and Bertrand and Marsh (2015), our study featured extensive direct observations of teacher data-interpretation meetings, supplemented by both individual and group teacher interviews. Specifically, we conducted more than 44 hours of observations of six grade-level teams of teachers who met, under a district initiative, to collectively analyze and make decisions from interim student assessments. In these grade-level meetings, we observed how and when explanations of student performance unfolded within teachers’ broader discussions of curriculum, instruction, students’ data, and educational mandates. Thus, this study presents insights into the ways in which teachers collectively connect data to instruction, and to other factors, in regularly occurring grade-level meetings, which are increasingly a part of K–12 teachers’ practice (Coburn & Turner, 2011; Little, 2012). Individual and group interviews with members of these teacher teams provided further opportunities to understand teachers’ sensemaking of the data use process and its contributions to both instruction and student learning.

In addition, we investigated a slightly different phenomenon than previous studies of teachers’ views of the connection between their instruction and student performance. Bertrand and Marsh (2015) studied teachers’ attributions, and Oláh et al. (2010) studied teachers’ diagnoses of students’ performance data. The concepts of attributions and diagnoses denote a systematic process wherein teachers thoughtfully attempt to pinpoint how and why students performed in a particular way on an assessment. Further, in the broader literature on attribution theory, a study of teachers’ attributions would examine factors like (a) whether the identified cause for students’ performance was internal (e.g., the child’s intellect) or external (e.g., the difficulty of the task); (b) the extent to which teachers have control over the identified cause of students’ performance; and (c) the perceived stability or the consistency of the cause of student performance over time (Bertrand & Marsh, 2015; Cooper & Burger, 1980; Jager & Denessen, 2015; Weiner, 2010).

In this study, the character of what we observed was not congruent with attribution theory. For example, we did not observe teachers attempting to systematically identify root causes of student performance. We also did
not observe teachers discussing if explanations of student performance were stable or if they could influence student performance. Instead, we observed teachers spontaneously calling out reasons or justifications for student performance that were at times only tangentially related to the data. We selected the term *explanations* because this, more appropriately than the term *attributions*, represents our observations of teachers brainstorming ideas for why students performed in particular ways. *Explanations* in this study thus refers broadly to teacher accounts or justifications for observed student performance rather than the analytic attribution of this performance to named causes.

**METHODS**

In this data use study, we employed a case study approach to better understand the relatively new policy-and-practice expectation that teachers examine, interpret, and then make instructional decisions from student performance data. This expectation is commonly labeled data-driven decision making. As Stake (1995) argued, case study is not a method but rather a selection of what to study. Given the lack of data on how teachers actually enact DDDM, we aimed to “thoroughly understand” how a sample of practicing teachers directly engage with data in their workplace (Stake, 1995, p. 9).

Previous research on teacher data use has most often relied on interview and survey methods, which capture individual teacher perspectives, and perhaps reflections, on data use in contrast to teachers’ actual, real-time engagement with data (Coburn & Turner, 2012; Little 2012). To capture teachers’ real-time engagement with data, our primary method was observing teachers as they engaged with student performance data during their grade-level team meetings over the course of one academic year. We also interviewed teachers, individually and in grade-level groups, to enhance our understanding of what occurred in these team meetings. Individual interviews \( n = 16 \) were conducted during the school year. Six end-of-study interviews were conducted, one with each grade-level team (see Table 1).
Table 1. Amount of Time Spent on Data Collection by School

<table>
<thead>
<tr>
<th>School</th>
<th>Observation (35 min on avg.)</th>
<th>Teacher Interviews (47 min on avg.)</th>
<th>End-of-Study Team Conversation (92 min on avg.)</th>
<th>Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riverdale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>16h 22m</td>
</tr>
<tr>
<td>Grade 5</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>13h 53m</td>
</tr>
<tr>
<td>Greenbrook</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>11h 41m</td>
</tr>
<tr>
<td>Grade 4</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>9h 21m</td>
</tr>
<tr>
<td>Clearview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>8h 39m</td>
</tr>
<tr>
<td>Grade 5</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>8h 53m</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>16</td>
<td>6</td>
<td>68h 49m</td>
</tr>
</tbody>
</table>

As suggested in the literature on supports for teacher data use (Marsh, 2012; Means et al., 2011; Schildkamp & Poortman, 2015), and as mandated by the school district, the teachers in this study practiced data use in grade-level teams. We observed nearly all the meetings of each of the six teams in our study over one school year. In our observations, we focused on the team as our unit of analysis, which is different from other researchers of teams (see, for example, Schildkamp & Poortman, 2015), who studied the interactions among individual team members. The teams thus constituted the cases in this case study research endeavor. This decision to study the team as a unit was partially influenced by teacher participants, who were increasingly suspect of being individually evaluated and held accountable for students’ performance data. Further, because teachers are often required to work in teams given the research that suggests teams are supportive (Marsh, 2012; Means et al., 2011; Schildkamp & Poortman, 2015), we wanted to provide an analysis of the conversations and sensemaking that occurred at the team level.

SCHOOL SITES AND DDDM CONTEXT

We recruited six grade-level teacher teams across three elementary school sites (two teams per school), given the pseudonyms Riverdale, Greenbrook, and Clearview. Each school was purposefully selected to serve a diverse student population, specifically with variation in students’ race and ethnicity, socioeconomic status, and achievement patterns (see Table 2). We expected that such diversity could manifest in distinct patterns of teacher sensemaking of performance data for different kinds of student learners.
Table 2. School Site Demographics

<table>
<thead>
<tr>
<th></th>
<th>Clearview (n = 424)</th>
<th>Greenbrook (n = 470)</th>
<th>River Dale (n = 365)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Income</td>
<td>53</td>
<td>81</td>
<td>59</td>
</tr>
<tr>
<td>African American</td>
<td>23</td>
<td>53</td>
<td>49</td>
</tr>
<tr>
<td>White</td>
<td>43</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Hispanic</td>
<td>12</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Multiracial</td>
<td>9</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Meet state performance standards in math and ELA</td>
<td>64</td>
<td>47</td>
<td>58</td>
</tr>
</tbody>
</table>

The teacher teams were purposefully selected from Grades 3–5, primarily on criteria related to expected information richness. Elementary-level classroom teachers know their students quite well compared with junior high and high school teachers. Poor student performance and inadequate or unsuccessful educational responses to it can be more consequential for upper elementary school students, compared with students in the primary grades. Specifically, student behavior and learning habits get more difficult to change with age. In addition, we recruited teams who had at least one year of experience of meeting regularly to analyze and make instructional decisions from student performance data. We expected experienced teams to have mastered the process and steps involved in data interpretation and use, so our observations could concentrate on data use themselves rather than learning about data-use activities.

As suggested in Earl (2009) and Farrell and Marsh (2016), teacher data-use routines, including who sets the agenda for DDDM and the type of student data examined, influence the way in which educators use data to inform their practice. In this study, the school district required all three schools in this sample to examine student performance data in grade-level teams. Teachers were mandated to meet in grade-level teams at least bi-weekly for a 30-minute block and quarterly for a 90-minute block to collectively examine student performance data. Each school site implemented this mandate differently.

At Greenbrook, the school administrator led the staff in examining student performance data for the sole purpose of making Response to Intervention (RtI) decisions. Essentially, student performance data were used to monitor student performance and intervene if students were not doing well on standardized assessments. At this school site, student data as a part of the RtI process primarily informed decisions about which students needed academic interventions and/or referrals to special
education. When making decisions based on data, teachers exclusively examined standardized assessment data from one district-mandated assessment. In one year of observations, we never observed discussions of teachers’ classroom-based assessment data. Further, the teachers’ role in these meetings was relatively limited compared with other school sites. Teachers predominantly responded to specific administrator-posed questions about whether students should be referred for special educator or placed in particular academic interventions.

At Clearview, the school administrator also led DDDM meetings. The administrator engaged educators in using student data to make RtI decisions in addition to broader discussions about how student data could inform instructional decisions, schoolwide polices, and initiatives aimed at dismantling the school’s achievement gap. At these meetings, teachers brought classroom-based assessment data, and the principal brought grade-level-wide data from multiple district-mandated standardized assessments. Teachers were invited to voice their perspectives on a variety of topics related to student performance data, yet the school leader posed the questions and steered the conversation.

Riverdale teachers had the most active role in DDDM conversations. Although instructional leaders were often present, teachers contributed most frequently to the conversation. Unlike other school sites, teachers always brought their own student performance data to the meeting. This included standardized assessment data from mandated districtwide assessments and standardized data from grade-level-wide assessments and classroom-based assessment data. Teachers set diverse agendas for these meetings, which ranged from entire meetings spent discussing one or two students’ performance to logistical issues with accessing student data. Of note, Riverdale was the only school where some educators had received intensive professional development on DDDM. Further, as a part of their DDDM practice, school staff had routine schoolwide conversations about student performance data, student behavior data, and how the school could better support students of color.

Despite differences in the assessment that data teachers examined, their level of autonomy when engaged with these data, and their aims for DDDM, we observed multiple similarities across school sites. First, most teachers indicated that if given the choice, they probably would have used grade-level meetings for a different purpose than examining student performance data. Across school sites, in individual and group meetings, the vast majority of teachers expressed frustration with the time spent on discussion student data. Second, many teachers expressed concerns for the assessment process, mainly that administering assessments was frequently taking away time from instruction. Further, most
teachers indicated that the information gleaned from these assessments they already had from working with students on a daily basis. Finally, across sites, DDDM was one of many mandates that teachers were attempting to satisfy. All teachers were implementing a new districtwide curriculum, partaking in a new, time-consuming teacher evaluation process, and experiencing a plethora of other new policies fostered by the state winning the Race to the Top competition for federal funds. Teachers often described DDDM in relation to their numerous other mandated responsibilities. In sum, teachers practiced DDDM differently across school sites, yet most teachers across our sample did not perceive their practice of DDDM as particularly meaningful.

DATA COLLECTION

From September 2013 to May 2015, five researchers observed 76 grade-level team meetings where participating teachers and other team members discussed students’ performance data. Again, we observed each team for one academic year, and we conducted individual and group teacher interviews as indicated earlier and displayed in Table 1. The team observations that constitute the database for this article are described in detail in the following section. The individual and group interviews are also further described next.

Collection of Observation Data

Grade-level teams of teachers met with varying frequencies, ranging from weekly to monthly, and we observed almost the entire number of team meetings that occurred during the study period. Across schools and grade-level teams, each team meeting occurred in a private conference room, away from students and other school staff. The topics of discussion at team meetings varied, and researchers observed the entire meeting even if student performance data were not the focus. A single researcher observed each meeting, took detailed field notes, and audio-recorded teachers’ conversations to assist the researcher in preparing the field notes, which constitute the primary data source for this article.

Drawing on the data-use framework by Coburn and Turner (2011), we systematically documented three aspects of teachers’ data use conversations in our field notes (see Appendix A for field note guide). First, from Coburn and Turner (2011), we documented when teachers “noticed” or focused their attention on particular data points during meetings, and we recorded the specific data points that were discussed by participants in our study. Second, Coburn and Turner (2011) also argued that teachers engage in “sensemaking,” defined as the process by which teachers relate
quantitative data to what they know about instruction, their students, and more. We attempted to capture teachers' sensemaking in our field notes by transcribing particular quotes or conversations when teachers interpreted data. Finally, per the aim of our broader research study, we documented in detail in our field notes all instances when student characteristics were invoked in data talk conversations. The corpus of data for this article thus consisted of 76 field notes, averaging five single-spaced pages in length.

**Interviews**

To supplement our understanding of what was occurring in data-use meetings, we conducted individual interviews with each participating teacher. The primary purpose of the interview was to pursue better understanding of how each teacher makes sense of the “data-use activities” observed during the grade level meetings. Specifically:

- How does the teacher make sense of the various data sets that he or she is asked to use?
- What supports has this teacher received for making sense of students’ performance data?
- In what ways do data contribute to teaching effectiveness for this teacher?
- In what ways do student characteristics influence or contribute to teacher data use?

Five different researchers used a semistructured interview guide to conduct one-to-one interviews with the teachers from each grade-level teacher team (see Appendix B for full interview protocol). These interviews were audio-recorded and then transcribed for analysis.

We also conducted an end-of-study group interview with each team, during which the researchers invited teachers to discuss preliminary study findings in an open-ended conversation that lasted from 60 to 120 minutes. For these group conversations, we shared a summary of preliminary study findings (a two-page handout, described next) and invited teachers to comment freely on these findings. We audio-recorded these conversations and created detailed field notes. Each field note included (a) instances when participants agreed, disagreed, or clarified the research team’s preliminary findings; (b) new insights on factors that influenced teachers’ data use at each school site; and (c) teachers’ wishes for data use, including the type of data they wished to examine and how they would like to use these data.
DATA ANALYSIS

At the early stages of data analysis, the research team reviewed all field notes to identify what we termed “data talk episodes.” These episodes were defined as instances in teachers’ conversations when they discussed a particular piece of data, either an individual student’s score or a group of students who scored in the same performance bracket, such as the following example.

Teachers 1 and 2 look at a spreadsheet, which displays students’ reading scores on multiple standardized assessments. Teacher 1 identifies a student who has a stronger score on a reading fluency test than on a reading comprehension test. Teacher 1 explains that when she evaluated his reading comprehension test, the student received a low reading reflection score because his response was not in alignment with the standardized evaluation rubric. Teacher 1 further explained that this student is “thinking outside the box . . . the story is too simple for his interpretation . . . an example of where a smart kid doesn’t match the rubric.”

Using an adaptation of Coburn and Turner (2011) “process of data use,” we dissected episodes of data talk into instances when teacher teams “noticed, interpreted, or constructed implications for data use” (Coburn & Turner, 2011, p. 176), such as the following example.

*Data Noticed:* Teacher 1 noticed how a particular student scored higher on his fluency assessment than on an assessment that emphasized reading comprehension.

*Data Interpretation:* No discussion of interpreting students’ quantitative scores.

*Explanation of Student Performance:* Teacher 1 explained the discrepancy in her student’s score as a “smart kid not matching the rubric.”

*Implications for Instruction:* None discussed

After dissecting teachers’ data talk episodes into each step of the process of data use, we looked for trends in each category. For example, we looked for trends across all data talk episodes in the category of teachers’ explanations of student performance. These trends comprised the two-page handout of preliminary findings shared during the end-of-study conversations with each grade-level team. Displayed next is the section of this handout, which focused on teachers’ explanations of student performance.
Teachers seldom related student performance to their own teaching (planning, preparation, instruction, classroom environment).

Teachers sometimes related student performance to the quality and usefulness of the assessment.

Teachers often related student performance to parent involvement, mother’s parenting, or other aspects about students such as motivation, attitude, language proficiency, reading stamina, self-esteem, engagement, or focus.

Overall, in the end-of-study team conversations, teachers confirmed that their grade-level team invoked particular explanations of student performance more consistently than others. We then used these initial trends as a springboard to a deeper analysis that aimed to identify categories of explanations of student performance and patterns for when those explanations were invoked.

Using the preliminary findings on teachers’ explanations of student performance as a guide, two researchers worked collaboratively to code teachers’ explanations of student performance across grade-level meetings, interviews, and the end-of-study conversations. Specifically, we coded for instances when teachers related students’ performance to (a) teachers’ instruction, (b) the quality or character of the assessment, and (c) student characteristics. These results were shared with the full research team on multiple occasions to clarify and revise the coding scheme. For example, during the coding process, the team decided the category of student characteristics was too diverse to be defensibly deemed one category. This resulted in new codes (students’ behavior, a suspected or established underlying condition of the student, and students’ home life), and the category of student characteristics became three distinct categories. Appendix C offers multiple examples of how we coded for teachers’ explanations of students’ performance.

Once the coding scheme was established, two researchers recoded a portion of field notes. Two different researchers also used the established coding scheme to code the same field notes, and we compared results to ensure consistency and to clarify our understanding of the coding scheme. Then two researchers completed coding all field notes and transcripts.

Based on our findings, we knew that particular explanations were invoked more readily than others, and often these trends were present across all six grade-level teams. To capture how particular categories of explanations were invoked more than others, we calculated an average for each category of explanation (see Table 3). For example, we identified all instances across observation and interview data when teachers invoked instruction as an explanation of student performance \((n = 25)\) and divided
this by the total number of explanations \((n = 168)\). We also calculated averages for each grade-level team and school site, which are described in the findings section. These averages offer insights into the extent to which particular categories of explanations were invoked more frequently than others. These results further complemented our more substantive interpretations and findings.

**FINDINGS**

Across grade-level teams and schools, we found that teachers’ explanations of student performance could defensibly be grouped into five categories, as presented in Table 3.

**Table 3. Definitions and Examples of Categories of Explanations of Low Student Performance**

<table>
<thead>
<tr>
<th>Explanation Categories</th>
<th>Description</th>
<th>Example</th>
<th>Percentage of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Behavioral characteristics of the student</td>
<td>Instances when teachers thought a student’s actions, attitudes, or behavior during class or when taking an assessment impacted the student’s test score</td>
<td>A teacher thinks a group of students did not do well on an assessment because they regularly do not pay attention in class.</td>
<td>32%</td>
</tr>
<tr>
<td>B. Mismatch between students and the assessment</td>
<td>Instances when teachers thought the format, content, or features of an assessment were inappropriate for a particular student</td>
<td>A teacher explains that English language learners in her class did not do well because the assessment was an oral, English fluency assessment.</td>
<td>21%</td>
</tr>
<tr>
<td>C. Suspected or established underlying condition of the student</td>
<td>Instances when teachers suspected students had a physical, mental, or emotional issue (enduring or temporary) that impacted students’ test scores</td>
<td>A teacher states to her colleagues that she thinks a low-performing student may have dyslexia.</td>
<td>18%</td>
</tr>
<tr>
<td>D. Students’ home life</td>
<td>Instances when teachers thought students’ parents, home environment, and/or life outside of school impacted students’ test scores</td>
<td>A teacher thinks a group of students did not do well on a test because the students do not read at home.</td>
<td>13%</td>
</tr>
<tr>
<td>E. Teachers’ instruction</td>
<td>Instances when teachers thought their lesson planning, instructional delivery, or curricula impacted students’ test scores</td>
<td>A teacher thinks students did not do well on a particular part of a test because she had not taught the concept tested in that section.</td>
<td>15%</td>
</tr>
</tbody>
</table>
Categories A, B, C, and D relate to teachers invoking student characteristics to explain student performance, and Category E contains instances of teachers connecting student performance to teachers’ instruction. We separated student characteristics into four distinct categories to better understand which particular student characteristics (or teachers’ perceptions of students) were invoked as reasons for student performance. For example, a teacher identifying a student’s classroom behavior as a root cause of low performance is distinguishable from a teacher suspecting that a student has a learning disability. Further, by having four distinct categories of explanations that relate to student characteristics, we were able to identify particular patterns in which student characteristics were most readily invoked as factors for student performance.

Across all six teams, we also found that teachers primarily discussed and attempted to identify reasons that students’ scores were low or unfavorable, rather than reasons for acceptable or strong performance. Teachers often discussed why, for example, students were in the lowest 10 percentiles on a nationally normed test or why a group of students scored lower than the rest of the class. Teachers rarely considered why students scored well or why test scores increased. Of all instances in which teachers offered explanations for student performance, 75% of those explanations related to unfavorable performance. Of note, at Greenbrook, teachers almost exclusively discussed unfavorable performance and only once offered an explanation of positive student performance. Across school sites, this emphasis on unfavorable student performance was relatively surprising because we expected teachers to consider possible root causes of high performance in order to learn from these instances. Yet, teachers’ discussions focused almost exclusively on the lowest performing students, and, correspondingly, we most frequently heard teachers seek to explain low or unfavorable student test scores.

**CATEGORY A. BEHAVIORAL CHARACTERISTICS OF THE STUDENT**

This category included all instances when teachers referenced students’ actions, attitudes, and/or other observable attributes of a child as a factor in students’ performance. When examining all explanations for student performance across school sites, teachers invoked this category of explanation more than others (32% of all explanations). When examining each school site, Clearview and Riverdale (44% and 37% of all explanations, respectively) invoked this type of explanation of performance more often than Greenbrook (16% of all explanations).

When using students’ behavior as an explanation for low performance, teachers typically discussed individual students who were perceived as
“not trying” or students who were misbehaving in class. For example, at a grade-level meeting at Riverdale, a student’s low performance was explained by the statement, “[student name] was doing his usual lost in space and searching the heavens” (FN²). In a similar example at Clearview, a teacher explained a student’s low performance as a “blatant lack of attention” and “it’s about the student not taking ownership” (FN). At Greenbrook, a teacher responded to one student’s low score with the comment, “Oh that is not what he can do at all. That kid could be in fifth grade right now if he were more into it” (FN). These types of explanations—individual students’ lack of investment in school or lack of attention—were consistently offered as explanations for low student performance in grade level meetings.

Occasionally, teachers explained the low performance of groups of students by citing what they regarded as a specific behavior or characteristic of the group. For example, during a grade-level meeting at Clearview, a teacher examined the lowest scoring group of students and explained the group’s performance by stating, “A lot of students aren’t invested. They don’t care” (FN). In a similar example, when discussing the low performance of the majority of students on a specific math item, a teacher at Riverdale stated, “My kids struggle with this because they are lazy. Pure laziness” (FN). In another example, at a grade-level meeting at Greenbrook, two teachers “bet” that the group of students with low reading scores were the same students who had a high number of discipline infractions (FN). In these and other examples, groups of low-performing students were perceived as behavior problems or lazy, and these characteristics were offered as explanations of students’ poor performance.

Similar to the finding in Bertrand and Marsh (2015), teachers suggested that particular unfavorable characteristics of individuals or groups of students were related to students’ performance. Seemingly global attributes of the child, such as laziness, were consistently cited as the reason for low test scores. The overriding theme in this category was that students with negatively perceived characteristics were culpable for their poor performance.

CATEGORY B. MISMATCH BETWEEN STUDENTS AND THE ASSESSMENT

Across school sites (21% of all explanations), teachers cited a mismatch between the student and the assessment as a reason for low test scores. We defined a mismatch as an instance when teachers thought the type, format, content, or features of an assessment were inappropriate for a particular student. A mismatch was typically invoked to explain that a low
summative or overall score on an assessment was an inaccurate indicator of students’ academic capabilities, as illustrated in the following exchange between two teachers at Riverdale:

Teacher 1: I’m really concerned about [student name] because he is very bright and [goes over his data] he gets it, but it takes time. So on the [state test] it’s going to look like he doesn’t get it.

Teacher 2: Maybe you need to teach him to watch the clock and monitor himself with a clock.

Teacher 1: I think I’ll start with a timer as the first thing. I need to shoot off a note to his parents too because they’ll be concerned. They’re going to see he’s so below but it’s not that he’s below he’s just taking his time. (FN)

As illustrated in this exchange, when invoking the mismatch explanation, teachers asserted that students were doing better academically than what a particular test indicated. As one teacher at Clearview stated, “The test didn’t reflect who [student name] was” (FN). This statement that an assessment does not reflect who the child is is representative of the mismatch category. In this category of explanations, teachers are making the argument that a specific assessment did not appropriately align with the child’s abilities, identity, native language, or other valued characteristics of the child. For example, in the preceding excerpt, the teacher explained a student’s low performance as his tendency to “take his time,” which did not align with the required time restrictions on the state test.

This idea that particular characteristics of an assessment do not appropriately match attributes of a child is illustrated in two particular instances when teachers considered the child’s culture in relation to the assessment. First, a third-grade teacher at Riverdale noticed that one student did poorly on a particular skill related to a student “looking inward,” a trait she stated was typically associated with societies that emphasize the individual, as opposed to the community. She argued, “Culturally this is not what the child has learned to do” (FN) and explained his poor performance to this mismatch between cultural norms. In a similar example at Greenbrook, a teacher and an ELL specialist observed that particular students performed poorly on the most recent fluency assessment. In the quote that follows, one of these teachers commented that students struggled with the names of the characters in the reading passage.

The passages are all supposed to be the same ability level, but I mean today it had the name Andy. How many of our kids know the name Andy? I got a lot of variations on Andy like Annie, Ambie, . . . and that name was over and over and over again and you have to
mark it as an error every time they get it wrong so there were kids who read it five times so that is five errors right off the bat. (FN)

The ELL specialist agreed and added, “That is a cultural observation, too.” In these unique instances, these teachers identified what is referred to in the literature as a “cultural mismatch” (Ladson-Billings, 1995; Villegas, 1988) between the cultures of individual students and the cultural values embedded in particular assessments.

Although reference to students’ culture was relatively unique, teachers were often concerned that diverse students were being evaluated using the same, standardized assessment tool. In this category of explanations, teachers often advocated that standardized assessment tools were inaccurately evaluating the skills and knowledge of particular students. An example of this occurred in a grade level meeting at Greenbrook.

The principal questioned why a particular third-grade student, Student A was in “enrichment,” an instructional setting designed to challenge a small group of high-performing students. The principal questioned why Student A was in enrichment because this student had scored in the 10th percentile on a nationally normed fluency test. The teacher responded, “[Student A], really?” with a tone of surprise in her voice. The principal double checked the data printout and confirmed this student had scored poorly on the assessment. The teacher then advocated for the student to remain in enrichment because the fluency test privileged speed over comprehension and the student was a “slow reader” but the teacher knew that she comprehended text well. (FN)

In another example at Riverdale, a team of teachers noticed that one student did well on one reading assessment and did poorly on the other assessment. This student’s homeroom teacher explained the poor test score as a mismatch between the student and the assessment; essentially she argued that the student was “smarter” than the rubric that accompanied the assessment. She stated that the student was “thinking outside the box. . . example of where a smart kid doesn’t match the rubric” (FN).

To summarize, in this category of explanations, teachers argued that particular test scores were inaccurate indicators of students’ academic abilities. They explained the inaccuracy of these scores as a mismatch between characteristics of an assessment and characteristics of students. From noting students as being “too smart” for a prescriptive rubric to students’ cultural values differing from the values embedded in an assessment, teachers utilized their knowledge of students to make sense of student performance.
CATEGORY C. SUSPECTED OR ESTABLISHED UNDERLYING CONDITION OF STUDENTS

When teachers attempted to explain students’ low performance, 18% of the time, they referenced a student’s established or suspected disability. The term suspected denotes that teachers consistently explained low test scores by stating their perception that students may have a disability. One purpose of data use in this district was to identify students who may need supplemental services, such as academic interventions or admittance into special education, so this category of explanation is aligned with this particular purpose of data use. We heard throughout interviews and grade-level observations that student performance data were a critical factor in referring students for specialized services/special education. For example, at one school site, a fifth-grade teacher and principal described how a particular student had severe behavioral and academic struggles in the classroom. The student’s difficulties were so extreme that the staff had been unable to assess this student. Yet, without data from assessments, the principal and teacher were unable to refer this student for specialized services. The principal stated, “The data is not there,” and that was a “conundrum” for the principal and teacher who had no doubt that the student required specialized services (FN).

The emphasis on DDDM for special education was especially salient at Greenbrook. Here, teachers primarily discussed student performance data to comply with the district’s RtI policy, that is, to identify students who may need additional instructional support and/or specialized services. Correspondingly, teachers at this school site most frequently explained low student performance as a suspected disability (37% of all explanations at Greenbrook). For example,

A homeroom teacher and a reading specialist discussed a particular student who had a very low reading test score. The reading specialist stated that this student had received “all kinds of services.” The specialist asked the homeroom teacher, “is she just a slow reader?” The homeroom teacher responded, “No, I think she is dyslexic. It doesn’t matter what you do today; it doesn’t transfer to tomorrow. Every day is brand new, every time she looks at a word, it is like it is brand new.” (FN)

In this excerpt, the teachers discuss a student who has no individualized education plan (IEP); as the teachers stated, “I think she is dyslexic.” We observed multiple instances like this where teachers explained students’ low test scores with the belief or suspicion that a child had a disability. Following is a similar example from Greenbrook, when a teacher suggested that a student was autistic:
The reading interventionist called out, “Look at this” and points at a particular student’s score. The instructional coach responded, “He is a good reader” while the homeroom teacher interjected, “He doesn’t understand anything he is reading.” The instructional coach responded, “Yes, he is a word caller,” and the homeroom teacher responded, “That has autism written all over it.” The instructional coach agreed. (FN)

This tendency to explain students’ unfavorable performance with the suspicion that a student has a disability also occurred at Riverdale and Clearview. The following is an example from Riverdale, where teachers were discussing one student who scored significantly lower than the rest of the class.

Teacher 1: “I’m thinking he doesn’t understand most of what we’re saying.”

Teacher 2: “No, he understands the English we’re saying. He can’t do it.”

Teacher 1: “But in directions he always turns to someone and asks what to do.”

Teacher 2: “So my thought, and this is because I had social work and psychology, he can understand what we’re saying but he doesn’t understand it. He can’t comprehend it.” (FN)

The teachers and instructional coach in the grade-level meeting continue to ponder if this student struggles academically because he has a disability or because he is an English language learner. We observed multiple instances in this category where teachers considered both students’ learning abilities and language abilities.

In this category of explanations, we rarely observed teachers reference a confirmed underlying condition, such as a student who had an IEP for a physical and/or a learning disability. In a few rare instances, we heard teachers relate performance to the specified goals or accommodations described on an IEP (FNs from all three schools). Even so, overall, the predominant trend in this category was teachers explaining student performance with the suspicion or perception that a student had an underlying condition that impacted his or her capacity to do well on a particular assessment.
CATEGORY D. STUDENTS’ HOME LIFE

The final type of explanation related to student characteristics was students’ home life. This category included instances when teachers identified students’ parents, home environment, or life outside of school as the reason for students’ test scores. Across schools, this type of explanation occurred approximately 13% of the time, almost as frequently as instruction. Like instruction, this category of explanation varied among teams. For example, at Clearview, one grade-level team referenced students’ home life more often (26% of all explanations) than their colleagues in a different grade-level team (8% of all explanations).

When using students’ home life as an explanation of low performance, teachers often said that students lacked support or resources at home. For example, a teacher at Clearview explained, “If they [students] don’t know where they are going to get their next meal from or are worried they are going to get evicted, homework isn’t the top thing on your mind.” Another teacher added that some students are the oldest child in their family and are put in charge of their younger siblings while their parents work an evening or night shift (FN). Later in the same conversation, a teacher added that the parents’ level of education was also a factor (FN).

During an interview with a fourth-grade teacher at Greenbrook, the teacher gave an example of how home life resulted in poor performance for one of her students.

One of my students, her mom is incarcerated and I don’t know where dad is in the picture and her grandma is taking care of her, however her grandma is battling cancer right now and going in and out of chemo and from what I can tell she’s just exhausted and there’s just not a whole lot of supervision after school. She’s far behind and she gets lost in whole group instruction. I’m constantly aware of her and trying to keep her on the same page but she’s just so far off, like she’ll get up for a second and start spinning around or just doing different things. I just try to ignore it and just to get through the lesson, because at the end of the day I still have 21 other students that are expected to be successful. So it’s a lot. (I)³

Teachers also related students’ positive performance to students’ home life. The following exchange took place during a third grade meeting at Riverdale:

Teacher 1: Wow, look at those scores! That’s wonderful!

Teacher 2: That’s what you call involvement. You can see even right here [looks at her chart of test scores]. Every person who is meeting [goals] right now has an involved parent (FN).
Similarly, a teacher at Greenbrook related the positive performance of her own students and the lower performance of other students to their home lives. “My students do well. A lot of the kids at this school don’t do well on the math test. I think that goes back to being a kid that got chances in life, was well fed, was supported, and had books at home” (I).

To summarize, when home life was used as an explanation, teachers cited parental involvement and resources at home as possible reasons for low or high student performance. A lack of resources and parent involvement was cited as a reason for poor performance, whereas the presence of those factors was seen as a reason for positive performance.

**CATEGORY E. TEACHERS’ INSTRUCTION**

This category represents the only instances when students’ performance was explained by teachers’ practice. Across schools, instruction accounted for 15% of all explanations. The use of this category of explanation varied widely among schools and grade-level teams. Teachers at Greenbrook only related students’ performance to instruction 6% of the time, whereas at Clearview, when teachers offered explanations for student performance, they referenced instruction 16% of the time. Riverdale had a significant grade-level difference between one grade-level team (27% of all explanations) and another (14% of all explanations). We also observed that teachers more readily related positive student performance to instruction (28% of all explanations of positive student performance).

Given that some literature on DDDM suggests that teachers ought to use student performance data to reflect on and evaluate their prior instruction (Duncan, 2009; Mandinach, 2012; Means et al., 2010; Moody & Dede, 2008), we specifically attempted to identify these types of occurrences. In alignment with the literature, at times, teachers’ explanations did include teachers’ reflection on the effectiveness of instruction of specific skills. For example, a third-grade teacher at Riverdale reflected, “I didn’t do a good job teaching that” when she saw the scores of a posttest indicating that only four students had mastered a particular skill (FN). In another example, in the excerpt that follows, a teacher at Clearview reflected that students did poorly on comparing two stories because the grade-level team had not taught students how to interpret.

A teacher stated that overall third grade students did a poor job on the writing prompt because they cannot compare two stories. She stated, “We keep testing but we don’t ever show them how to interpret. We need a break from all of this weekly testing so we can reteach.” The other third grade teachers all nod and agree. (FN)
In a third example, a teacher at Riverdale used students’ performance to pose an instructional question to her grade-level team. On showing a student’s lack of progress on a written assessment, she stated, “These data reflect my current question of ‘how do I help my students become better writers?’” (FN). In all three of these examples, teachers were using students’ performance data to reflect on their instructional decisions and actions.

Although we did observe teachers using data to reflect on their prior instruction, these types of occurrences account for only roughly half of the excerpts in this category. We also observed instances where teachers explained performance by citing instruction, but not their instruction. This is illustrated by a fifth-grade teacher at Riverdale whose student math scores were low. On discussing this unfavorable performance in teacher team meeting, she stated that these low scores were a consequence of the subpar instruction provided by their previous teacher. As she put it, “We all know she didn’t teach,” referring to the students’ fourth-grade teacher (FN). Another example of this involved teachers relating low performance to missed instruction due to students being absent and therefore not learning content on the assessment (FNs from all three schools). In a third and final example, a principal asked teachers at Clearview to consider whether aspects of their instruction negatively impacted achievement. In considering their instruction, teachers mainly focused on components of the district-mandated math and reading curriculum. Teachers commented that the pace of the reading curriculum was “quite rapid” and that students “don’t like” it (FN). For the math curriculum, teachers commented that the online resources only worked “sporadically,” and the professional development sessions on this curriculum were “not very useful.” Again, teachers were explaining low performance by citing instruction, but they were referring to aspects of instruction that they did not control.

DISCUSSION

Our findings are consistent with the findings of Bertrand and Marsh (2015) and Oláh et al. (2010). What is apparent across all three studies is that teachers consistently connected students’ performance to particular student characteristics, notably both verifiable and unverifiable student characteristics. Bertrand and Marsh (2015) reported stereotypes of particular groups of students: Namely, students with first languages other than English and students with an IEP were at times invoked as the explanation of low student performance. Oláh et al. (2010) also cited limited English proficiency, low levels of motivation, and other student characteristics as attributions teachers offered for student performance. Like Bertrand and
Marsh (2015) and Oláh et al. (2010), we found that teachers considered various characteristics of students as the problem that fostered low-achievement. This collective finding departs from both scholarly and political discussions on data use (Duncan, 2009; Mandinach, 2012; Means et al., 2010); we found that in practice, teachers do not consistently connect student performance, particularly unfavorable student performance, to their instructional actions and decisions.

This study departs from the findings of Bertrand and Marsh (2015) and Oláh et al. (2010) in three ways. First, by examining four distinct categories of student characteristics, we found that specific categories of student characteristics operated differently as explanations of student performance. For example, when teachers thought that student characteristics were misaligned with an assessment, low test scores were seen as inaccurate representations of students’ abilities. In these instances, the problem identified was a misalignment between the student and the test. Yet, in instances when teachers were relating low test scores to students’ behavior or actions, the problem identified was typically the student. This is an important distinction for how teachers see the “problem” that is causing low test scores.

Second, we identified a category of explanation for student performance that does not appear in the current literature: teachers relating student performance to the suspicion that students have an underlying condition, such as a disability. The suspicion that students may have a condition that impedes their academic performance is distinct from teachers connecting student performance to a diagnosed disability. We believe this explanation is related to the widely implemented policy of RtI. For RtI, teachers examine student performance data to identify and respond to students who may test below their peers. RtI is intended to provide low-performing students with academic or behavior supports. If students do not respond to these additional supports as evidenced by an increase in test scores, then students are referred for special education (Fuchs & Fuchs, 2006). In other words, RtI creates an instructional space between general education and special education where low-performing students receive extra tutoring or academic interventions without being a part of special education. This gray area between general and special education may have fostered this new category of explanation in which teachers related low student performance to the suspicion of a disability. The influence of policies like RtI and DDDM on teachers’ interpretations of student performance is an area that requires further research.

Third, our findings suggest that teachers connect student performance to student characteristics more often than to instruction. Bertrand and Marsh (2015) found the opposite, or that teachers related student
performance to instruction more frequently than to student characteristics, although student characteristics were the second most frequent type of attribution. Oláh et al. (2010) did not report specifics on frequency but did state that “by far the most common diagnosis” was the procedural category, or students’ difficulty with mathematical step-by-step processes for answering questions (p. 238). One possible factor for the difference in frequency among studies could relate to the type of assessments and student performance data teachers examined across studies. Another possible factor is the difference in methods used across studies. Further research examining both possibilities could offer additional insights into how readily and under what circumstances teachers relate their instruction to student performance data.

CONCLUSION

Although a major finding of this study was that teachers most frequently explained low performance by citing characteristics of the student, we do not want the takeaway message to be that we should blame teachers for blaming students. Instead, we suggest the need to distinguish between helpful and harmful claims that teachers make about students and their academic performance. Our findings denote a need to differentiate between teachers’ claims about students that are verifiable and those that are subjective, particularly negative, opinions about children. We suggest that this current political emphasis on “evidence-based” teaching provides an opportunity to challenge subjective, negative claims about students and in turn place a greater value on teachers’ verifiable claims about students and their academic needs. In other words, we suggest that teachers’ knowledge of students is important in instructional decision making, but teachers’ low expectations and/or stereotypes of students are not.

Although this assertion seems like common sense, the current emphasis on students’ performance data and evidence-based teaching often devalues all of teachers’ insights on students. Particular policies and research on DDDM suggest that teachers’ personal knowledge of students is irrelevant in instructional decisions (Duncan, 2009; Mandinach, 2012; Means et al., 2010). Teachers’ “gut feeling” and knowledge of students are often discounted as less valuable and potentially dangerous in instructional decision making (Mandinach, 2012, p. 71). Granted, particular bodies of research rightfully support this stance that teachers’ perceptions and judgment can be potentially harmful in instructional decision making. A substantial body of research on teachers’ attributions and teachers’ sense of responsibility for student learning suggests that teachers’ ideologies and bias can foster inequitable learning environments
(Diamond & Spillane, 2004; Harris, 2012; Johnson & La Salle, 2010; Park, Daly, & Guerra, 2013). Educational research clearly supports the claim that students of color, students with disabilities, and English language learners often lose when educational and instructional decisions are made according to educators’ perceptions (Booher-Jennings, 2005; Boykin, 2000; McKown & Weinstein, 2008).

On the one hand, our study offers additional support for the notion that teachers’ judgment is not productive in instructional decision making. Like Bertrand and Marsh (2015), Barrett (2009), and multiple other studies, this study demonstrates that teachers may have a problematic tendency to perceive low performance as a product of a defective learner or one that is “lazy” or “inattentive.” We too suggest that this tendency is problematic, particularly considering that we found teachers related poor performance to negative student characteristics approximately a third of the time.

On the other hand, we propose that teachers’ use of student characteristics in instructional decision making is at times highly informative. Like the suggestion found in Mandinach and Gummer (2016) and the seminal work of Schulman (1987), our findings demonstrate how teachers’ knowledge of students is instrumental in decision making. Like Schulman’s (1987) concept of adaptation, teachers in this study “adapted” the meaning of student data by applying knowledge of the students to their interpretations of quantitative performance data. In this way, teachers in this study capitalized on valuable knowledge of students, their strengths, and their needs that contributed to them making sense of students’ low performance. For example, in the mismatch category of explanations, teachers used their knowledge of students to assert that students were smarter and more capable than the assessment results indicated. In this way, teachers used their knowledge of students to advocate for an assessment that appropriately accounted for the diversity of students.

Moving forward, we propose that in research and in practice, scholars and practitioners can more carefully distinguish between the valuable knowledge teachers contribute to instructional decision making and the problematic ways educators may see students’ poor performance as a product of students’ poor behavior. Instead of promoting students’ performance data over teachers’ judgment, we suggest a more thoughtful consideration of what diverse sources of evidence, including teachers’ knowledge of students, can support teachers’ instructional decision making.
NOTES

1. We are grateful to the Spencer Foundation for funding this research as a part of its strategic initiative on Data Use and Educational Improvement.

2. (FN) denotes a field note or data collected from a grade-level team observation.

3. (I) denotes data retrieved from an interview.
REFERENCES


APPENDIX A

Grade-Level Team Observation Field Note Guide

DESCRIPTORS

- School
- Team
- Date
- Time (start / stop)
- Observer
- Names / Titles of all others present + seating arrangement

DATA AT MEETING

Use the table below to describe the data used or referred to at the meeting. For each data set, describe the subject matter (e.g., math, reading), data source (e.g., benchmark assessment), type of data (e.g., individual student scores, item-level data), data display (e.g., printed spreadsheet, computer display). As possible, get a copy of the data (with identifiers deleted).

<table>
<thead>
<tr>
<th>Name</th>
<th>Subject Matter</th>
<th>Data Source</th>
<th>Type of Data</th>
<th>Data Display</th>
<th>Copy of data obtained?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SUMMARY

Provide a summary of the meeting, including whether there is a set agenda, purpose and topics of the meeting and brief summary of what happened.

NARRATIVE ACCOUNT

From notes and audio recording, prepare a descriptive, chronological account of the interactions and discourse of the meeting. Note who is speaking and for how long. Note what data are being discussed or what other topics are being discussed. Also record nonverbal interactions and gestures (e.g., raised eyebrows), as well as the “tone” of the meeting that you were able to observe. For especially relevant parts of the meeting, transcribe selected dialogue from the audio recording. Be inclusive of all possibly relevant conversation. Beyond a summary statement, no need to detail irrelevant conversations, e.g., on food or baseball.

Feel free to comment on any parts of your descriptive account with an observer comment [OC: comment].
METHODOLOGICAL COMMENTS

Comment on the quality of the data in this field note. Specify all concerns you have regarding data quality (e.g., you couldn’t hear well, the team was not really on task).

ANALYTIC, INTERPRETIVE COMMENTS

Comment on what was learned from this observation related to our study foci. Address the following questions:

1. Participant relationships and engagement
   How was the meeting conducted? Was there a clear leader who directed the conversation to specific issues or questions? Did everyone participate? What was the level of engagement among meeting participants?

2. Character and tenor of meeting
   How would you describe the overall character or tenor of the meeting, noting in particular silences, tensions, disruptions, as well as harmonies and connections? Were there disagreements or even arguments, and if so, over what?

3. Character and quality of data use conversations
   What was the character and quality of the data use conversations? What specifically did the team spend its time discussing—trying to understand a set of data, discussing the quality or accuracy or relevance of some data, including student characteristics (which ones?) as part of data interpretation, discussing instructional or other uses of the data, other? How would you characterize the quality of the data use conversation? What were the decisions that were being debated (e.g., what to teach, how to teach, how to manage a number of kids), and how did evidence play into that?

4. Student characteristic data
   Elaborate here on if and how student characteristic was discussed in this meeting. Which student characteristics were discussed? How did they pertain to individuals or groups of students? How were they discussed relative to discussions of student performance data and/or instructional decisions?

5. Politics of data use
   In what ways, if any, did the conversation reference the politics of data use, for example, requirements or expectations regarding data use imposed from the outside? How salient were such conversations as part of this data use meeting? How does the politics of the data
systems (e.g., what is valued in what is measured) show up? How does it define achievement and learning? It is discussed or just taken at face value?

6. Resolution of issues and topics
How well did the team come to some resolution of the issues and topics discussed? Which issues were resolved, and which were left on the table? What do you think both facilitated and hindered resolution?

7. Areas of further inquiry

8. Anything else of importance

9. Running record
APPENDIX B

Teacher Interview Guide

In a revised project schedule, individual teacher interviews will be conducted mid-year. The guide that follows is intended to be inclusive of all of the issues we would like to discuss with teachers. Interviewers are free to select those issues most relevant to each teacher. Expected time of interview, 60 minutes.

INTERVIEW GUIDE

Introductions, informed consent, permission to tape

1. [Ice-breaker, could also use some other comfortable question to get started]
   I am so pleased that we have this chance to talk a little bit. I always enjoy chatting with fellow teachers. And I am always curious about their motivations for going into teaching, which is such a difficult job. Could you share with me your initial motivations for going into teaching? [Have a wee conversation here]

2. I would also like to invite you to talk a little bit about the students in your classroom this year. I imagine that each class of students is unique. How would you describe this year’s class as a whole?
   • Now can you further describe some of the individual students in your current class, in whatever ways make most sense to you? No names or first names only.
   • And now, specifically related to learning and academic achievement, how would you characterize the profiles or clusters of your students this year?
     - Children who “stand out” in some particular way?
     - Children who you are concerned about?
     - Children who you are delighted to have in your class?
   • Anything else you might wish to add about the particular children in your class this year?

   Brief member check on questions 1 and 2

3. Now I would like to begin our conversation about your grade-level data use meetings. As you recall, these are the main focus of our research project. So, let’s begin.
   • Questions on a few carefully selected “issues” that seem most relevant to this teacher.
   • Possible issues and questions:
• Overall perceptions of data systems and the data they provide:
  • What is your overall view of the data that are provided by the various data systems in use at your school?
  • Which data systems are most useful to you? In what ways are they useful?
  • What do you know about your students from the data that you didn’t already know from your own observations and interactions with your students?
  • How would you describe the alignment of the data systems with the various curricula you are currently using, for example, the alignment of [name of assessment with [name of curriculum used at particular school]?

• Understandings of “data use”:
  • What do you understand as your specific responsibilities for data use from these data systems? How well do the data systems and the data they provide enable you to meet these responsibilities?
  • Are you sometimes surprised by the data for an individual child, for example, surprised that the child scores at a yellow level and not a green level, or vice versa? What do you do in that kind of situation? (Do you have stronger confidence in one or the other, and if so, for what reasons? Do you try to reconcile the data with your own observations and expectations? …)
  • What other kinds of data and information on your students contribute to your instructional decision making? When (in what kinds of situations or for what kinds of children) do you bring those data to bear?
  • More generally, how does your knowledge of the assets and challenges each of your students brings to school play a role in your data interpretation and use? What do you do with all of the other information you have on your students?

• Contributions of data to teaching effectiveness:
  • In your striving to be the best teacher you can be, what are you doing differently now that these data systems and data use are part of this district’s educational system?
  • How would you evaluate the value or contribution of the data now available to your overall teaching effectiveness?
  • What else is critical to you for realizing your aspirations to be the best teacher you can be?
• **Support** for data interpretation and use:
  • Data interpretation and use are difficult tasks. How well supported do you feel in accomplishing these tasks this year?
  • What do you perceive as the primary sources of support for your data use? And what are the main obstacles or challenges? (Probe for role of school climate/culture if not mentioned voluntarily.)
  • Please describe any relevant professional development activities that have helped you engage meaningfully in data use and interpretation.
  • In your view, what kinds of additional support would be especially useful to you in fulfilling your data interpretation and use responsibilities?
  • **Other** – what other issues and concerns related to district/school data systems initiatives are important to discuss?

**Member checks along the way**

**Wrap-up**
  • Do you have any questions for me?

*Thank yous…*
APPENDIX C

Sample Coding for Teachers’ Explanations of Student Performance

CODES

Students’ behavioral characteristics (B), mismatch between student and the assessment (M), suspected or established underlying condition of the student (S), students’ home life (H), teachers’ instruction (I)

Data Talk Episode 1

Teacher 1 says, “Wow Teacher 2, look at those scores! That’s wonderful!” Teacher 2 states, “Yup, that’s what you call involvement. Yup, that’s what that’s called. You can see even right here (looks at her standardized data displayed on a chart). Every student who is meeting [goals] right now has an involved parent (H).

Data Talk Episode 2

Looking at an entire group of students’ performance data, Teacher 1 stated that the students needed harder passages, although some of the students struggled. She stated that three girls stand out in terms of language processing issues. One girl lacks focus like her brother (everyone remembers the brother and sighs collectively) (B). The other two girls also have language processing problems (S).

Data Talk Episode 3

Teacher 1 asks, “What do we do with students who scored below the first performance bracket?” This teacher has identified 1 student who scored very low. The other teacher starts to say that they don’t have anyone in this category, and Teacher 1 responds, “Yes we do, his name is (student name.)” Teacher 2 (sounding frustrated with student), “Well, he didn’t do anything with all the time and all the support.” Teacher 2 discussed that they worked on this assignment for a long time and he would come in every day, “Like what are we doing?” The specialists commented how this student gave her attitude when she came in to help with this project. Teacher 1 stated that this is because he just doesn’t understand. The specialists responded, “He understood, he was being rude.” Teacher 1 starts here to advocate for this student and provides examples of how she truly believes that he does not understand what is going on. This assertion shifts the conversation to a discussion of this student needing help. They begin to discuss specific examples of him not following/
hearing/understanding specific instructions (S). Teacher 1 asserted that she needs help with this student and they continue to discuss his needs based on anecdotal observations. (OC: Throughout the conversation, she is really pushing to go from talking about this student to concrete steps for how to support this student.)

**Data Talk Episode 4**

Teacher 1 says she is worried about four other kids who are showing no growth. She tries to figure out why. She says one problem is the one student lies to her about doing reading (B). Teacher 1 continues that another child’s parents do not take it seriously (H). Instructional coach agrees.

**Data Talk Episode 5**

Teacher 1 adds, “Some of my high guys aren’t scoring as well as they used to.” Teacher 2 comments, “I personally think this test is too hard. Teacher 3, who took the test herself said, “I would pick this answer, but the answer was wrong.” Teacher 1 adds that he also took the test twice and got an 80% and a 90% (M).
MARGARET EVANS is an assistant professor at Illinois Wesleyan University. Dr. Evans supports the development of future K–12 teachers for social justice. Her research interests include teachers’ data-driven decision making and the ways in which educators are responsive to the needs of students living in poverty.

REBECCA M. TEASDALE is a doctoral student in educational psychology at the University of Illinois at Urbana-Champaign. Her research focuses on methodology for evaluating adult learning and informal science learning, with a particular emphasis on representing learners’ perspectives in the valuing process.

NORA GANNON-SLATER is the director of performance and analytics for Breakthrough Charter Schools, an urban charter school network located in Cleveland, Ohio. As part of her work, she works with education practitioners and policy makers to adopt research-based policies and practices, and she designs tools, resources, protocols, and routines to improve individual and organizational capacities to incorporate systematic evidence in decision-making processes. Nora has presented her work at local and national conferences as well as state education departments. Most recently, Nora coauthored a special issue on equity and data use in the Journal of Educational Administration (in press) with Amanda Datnow and Jennifer Greene.

PRIYA G. LA LONDE is assistant professor of teaching at Georgetown University. She studies data and research use, international comparisons of incentivist policy, and social justice education. Her current work examines how performance pay shapes teacher relationships and work in high-performing schools in Shanghai.

HOPE L. CRENSHAW is the training and program coordinator of Teen Health Mississippi/Mississippi First. Her research interests include professional development, parent engagement, equity, and evidence-based interventions.

JENNIFER C. GREENE is a professor of educational psychology at the University of Illinois, Urbana-Champaign. Greene’s work focuses on the intersection of social science methodology and social policy and aspires to be both methodologically innovative and socially responsible. Her research interests include democratic valuing in evaluation, and methodological innovation with accountability. Two recent publications in these domains are: Greene, J. C. (2016). Advancing equity: Cultivating an

THOMAS A. SCHWANDT is professor emeritus, Department of Educational Psychology, University of Illinois at Urbana Champaign. His scholarship is focused on theory of evaluation, and his most recent book is *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice* (Stanford University Press, 2015).