

Data-Based Decision-Making: Developing a Method for Capturing Teachers' Understanding of CBM Graphs

Christine A. Espin*, Miya Miura Wayman†, Stanley L. Deno, and Kristen L. McMaster

University of Minnesota

Mark de Rooij

Leiden University

In this special issue, we explore the decision-making aspect of *data-based decision-making*. The articles in the issue address a wide range of research questions, designs, methods, and analyses, but all focus on data-based decision-making for students with learning difficulties. In this first article, we introduce the topic of data-based decision-making and provide an overview of the special issue. We then describe a small, exploratory study designed to develop a method for studying teachers' understanding and interpretation of Curriculum-Based Measurement (CBM) graphs. Specifically, we examine whether think-alouds scored for coherence, specificity, reflectivity, and accuracy differentiate teachers with more or less understanding of CBM data. We conclude the article by discussing the importance of, and the need for, research on teachers' understanding, interpretation, and use of data for instructional decision-making.

EXAMINING TEACHERS' DATA-BASED DECISION-MAKING: INTRODUCTION TO SPECIAL ISSUE

The use of data for decision-making has become increasingly important at all levels of education (Mandinach, 2012). In a presidential address to the American Psychological Association, Mandinach (2012) described *data-driven decision-making* as an “essential component of education,” arguing that it was “no longer acceptable to simply use anecdotes, gut feelings, or opinions as the basis for decisions” (p. 71) in educational research or practice. Mandinach (2012) defined data-driven decision-making as the “systematic collection, analysis, examination, and interpretation of data to inform practice and policy in educational settings” (p. 71). This definition highlights one of the greatest challenges in data use in education: The data must not only be collected and analyzed—they must be *examined and interpreted*, and done so in a way to inform practice and policy. Yet the examination and interpretation of data to inform teaching and to improve student learning has been a challenge for educators in both general and special education (e.g., Black & Wiliam,

2005; Datnow & Hubbard, 2016; Mandinach & Gummer, 2013; Schildkamp, Ehren, & Lai, 2012; Stecker, L. Fuchs, & Fuchs, 2005; Young & Kim, 2010). That data use is a challenge should perhaps not be surprising, given what is known about human decision-making in general.

Human Decision-Making

The literature on human decision-making reveals that, in general, people have difficulty using data to make decisions (see, for example, Ariely, 2009; Baron, 2000; Koehler & Harvey, 2007). People often make what appear to be “irrational” decisions that do not take into account available data (e.g., Baron, 2000; Koehler & Harvey, 2007). Simon (1990) describes human decision-making as “bounded rationality,” reflecting the fact that the use of data in decision-making is bounded by human limitations such as short- and long-term memory restrictions or limitations in recognition time (see also Gigerenzer, 2007).

Research on decision-making in “real world” or naturalistic settings such as medicine, aviation, sports, economics, and fire-fighting often makes use of an expert/novice paradigm in which the decision-making processes used by experts are compared to processes used by novices (see Ericsson, Charness, Feltovich, & Hoffman, 2006; Wickens, Lee, Liu, & Gordon-Becker, 2004, chapter on decision-making). A few of the consistent findings to emerge from this line of research are that (1) experience and expertise are not the same; (2) knowledge of content matter is important to expertise; (3) expertise involves selective access to important or relevant information; and (4) experts not only know more, they know “different,” meaning that they build an integrated and coherent knowledge structure of the field (Ericsson, 2006a;

*Present address: Leiden University, 4A.25 Pieter de la Court Gebouw, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.

†Present address: 22W408 Birchwood Drive, Glen Ellyn, IL 60137 USA.

Note: While we were working on this article, our dear friend and colleague Stanley Deno passed away. We are so grateful that we had the opportunity to know and work with Stan. We will miss his insights, his humor, his care and concern for those around him, and his dedication to science. Stan, thank you for being such a wonderful mentor to so many. We're glad we got to share this “piece of bologna” with you. We will miss you. The field will miss you.

Requests for reprints should be sent to Christine Espin, Leiden University. Electronic inquiries should be sent to espinca@fsw.leidenuniv.nl.

Ericsson & Ward, 2007; Feltovich, Prietula, & Ericsson, 2006).

The expert/novice paradigm also has been applied to the study of teacher decision-making (e.g., see Berliner, 1989, 1994; Perkins, 2009; Sabers, Cushing, & Berliner, 1991; Sternberg & Horvath, 1995; Swanson, O'Connor, & Cooney, 1990; Tsui, 2003), and results have mirrored those of the expertise research. Compared to novices, expert teachers have more and better-organized domain-specific knowledge and are more efficient in decision-making, due in part to their integrated, coherent content knowledge (Sternberg & Horvath, 1995). Despite the fairly large body of research examining teacher decision-making, little attention has been devoted to decision-making for students with learning and behavioral difficulties (see exceptions, Coddington, Skowron, & Pace, 2005; DeStefano, Shriner, & Lloyd, 2001), yet these students present the greatest challenges to, and are most likely to be affected by, the decisions made by educators.

Decision-making for students with learning and behavioral difficulties often involves the use of ongoing progress data. Such data appear on a graph, introducing an additional factor in the decision-making process: Before using data to inform their instruction, teachers must understand and interpret progress graphs; yet people vary in their ability to understand and interpret graphs, often experiencing difficulty with interpreting even simple graphs (see Canham & Hegarty, 2010; Friel, Curcio & Bright, 2001; Garcia-Retamero, Okan, & Cokely, 2012; Kratochwill, Levin, Horner, & Swoboda, 2014; Shah & Freedman, 2011; Zacks & Tversky, 1999).

Fortunately, both the decision-making and graph-interpretation literature reveal that performance improves with training and development. Experts in fields such as medicine, aviation, sports, economics, firefighting, and teaching have learned to select the most important and relevant data from all the data available to them, and integrate them into a coherent knowledge structure that guides their decision-making (see Berliner, 1989, 1994; Ericsson et al., 2006; Perkins, 2009; Sternberg & Horvath, 1995). Related to graph interpretation, experts are better able to select relevant data from graphs, to integrate and interpret the relations between the data, and to draw inferences from those relations to solve problems or answer questions (Curcio, 1987; Friel et al., 2001). Curcio and Friel et al. describe these levels of graph interpretation as *reading the data*, *reading between the data*, and *reading beyond the data* (Curcio, 1987; Friel et al., 2001).

In sum, both the decision-making and graph-interpretation research suggest that it is erroneous to assume that teachers will effortlessly use data to make instructional decisions. There is a need for systematic and explicit examination of data-based decision-making processes, and of ways to improve such processes. Such research should focus on educators' ability to understand and interpret (graphed) data, and on their ability to use data to inform educational decisions.

Overview of Special Issue

In this special issue, we explore educators' ability to understand, interpret, and use data for instructional decision-making. The first three articles focus on the use of

Curriculum-Based Measurement (CBM) progress data (Deno, 1985), and examine teachers' understanding and interpretation of CBM graphs. The first (this) article describes the development of a method for studying teachers' understanding of CBM graphed data. In the two subsequent articles, this method is applied and expanded to examine preservice (Wagner et al.) and in-service (Bosch et al.) teachers' abilities to interpret CBM graphed data. The fourth and fifth articles focus on progress monitoring via systems other than CBM. The fourth article (Zeuch et al.) describes the development of a learning-progress-graphics test that can be used to assess teachers' abilities to interpret progress graphs, and describes the thinking processes used by teachers when reading progress graphs. The fifth article (Keuning et al.) takes a somewhat different tack, focusing on data use, and examines teacher- and school-related factors that affect implementation of data-based decision-making. The progress data in this last article are from a national, standardized student progress monitoring system.

The articles included in this issue address a wide range of research questions, designs, methods, and analyses, and present an international perspective, with research reported from three different countries. Despite this diversity, the articles address one common theme: Data-based decision-making for students with learning difficulties.

Data-Based Decision-Making in Special Education

Although in recent years, as a result of reforms such as No Child Left Behind, data-driven decision-making has taken on increased importance in general education (Datnow & Hubbard, 2016; Mandinach, 2012; Schildkamp et al., 2012; Young & Kim, 2010), in special education the concept of data-driven decision-making is not new. As early as 1977, Deno and Mirkin described an approach that had at its core the use of data for instructional decision-making. This approach, called Data-Based Program Modification (DBPM: Deno & Mirkin, 1977; L. Fuchs, Deno, & Mirkin, 1984), was defined as an empirically-oriented approach to individualizing educational plans for students with learning and behavior problems. DBPM involved the systematic use of data-based procedures to evaluate the effectiveness of alternative solutions for students who experienced difficulties in schools. An important assumption underlying DBPM was that one could not determine a priori whether a program of instruction or an intervention would be effective for an individual student; it was therefore necessary to treat instructional programs and interventions as instructional "hypotheses" that could be "empirically tested" before a decision could be made about their effectiveness for a given student (Deno & Mirkin, 1977, p. 11). This approach has come to be known as *data-based decision-making*.

In recent years, data-based decision-making has taken on new importance in special education with the implementation of schoolwide problem solving models such as Multi-Tiered Systems of Support (MTSS) and Response to Intervention (RTI). In these approaches, data are used to make decisions

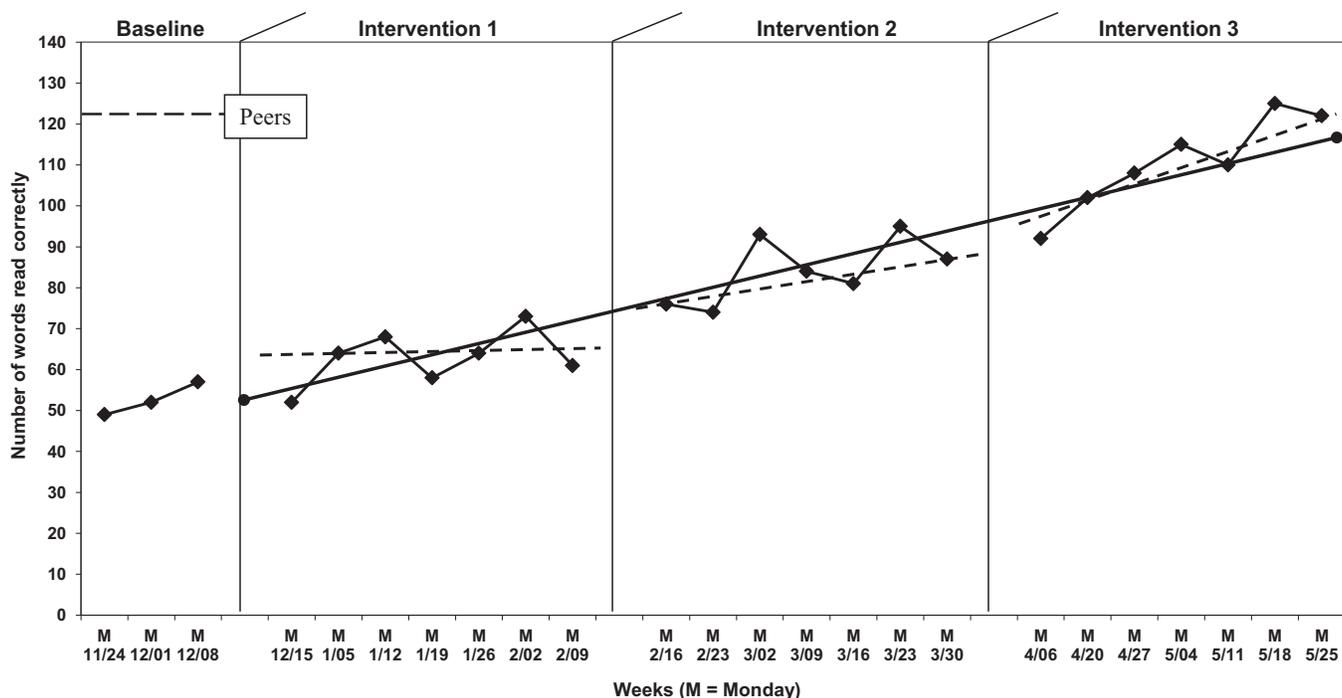


FIGURE 1 Sample CBM graph.

regarding the appropriateness and effectiveness of instruction within “tiers” of instruction (L. Fuchs & Vaughn, 2012; Vaughn & L. Fuchs, 2003). The intensity of instruction increases with each tier, with Tier 3 instruction being the most intensive and individualized.

Until recently, the majority of research and writing on MTSS/RTI approaches has focused on Tiers 1 and 2, with little attention directed toward Tier 3 (Berry Kuchle, Zumeta Edmonds, Danielson, Peterson, & Riley-Tillman, 2015). This is unfortunate because Tier 3 is the level where the most “specialized” instruction occurs—that is, the level where intensive, individualized interventions are designed and delivered for students with severe and persistent learning and behavior difficulties. In recent years, more attention has been devoted to Tier 3; see, for example, the special issue of *Learning Disabilities Research and Practice* devoted to Tier 3 interventions (volume 30, issue 4, 2015) and the National Center on Intensive Intervention (NCII; <http://www.intensiveintervention.org>). With the increasing interest in Tier 3 has come a reemergence of interest in the ideas originally articulated by Deno and Mirkin in Data-Based Program Modification (see, for example, Berry Kuchle et al., 2015). Foremost among these is the idea that interventions for individuals with severe and persistent learning and behavior difficulties should be treated as instructional hypotheses that need to be empirically tested to determine their effectiveness. Such “empirical testing” involves ongoing data collection to track an individual’s performance and progress under various instructional conditions. One system of measurement that is ideally suited to the empirical testing of interventions for individual students is *Curriculum-Based Measurement* (Deno, 1985, 2003).

Data-Based Decision-Making with CBM

In CBM, the progress of students with learning difficulties is monitored on a weekly basis, and the data are presented in graphs that display the student’s growth in an academic area over time (see Figure 1). When implementing CBM, teachers inspect the graphs on a regular basis to determine if the student’s instructional program is effective. Based on the growth rates displayed on the graph, the teacher makes a decision to either continue to collect data (growth is at the expected rate), to change instruction (growth is at a rate less than expected), or to raise the goal (growth is at a rate greater than expected). By systematically responding to the data with instructional and goal changes, teachers build effective instructional programs that lead to improvements in student performance (see review, Stecker et al., 2005). Unfortunately, research reveals that teachers often have difficulty responding to the data for instructional decision-making (Stecker et al., 2005).

In the 1980s and 1990s, L. Fuchs and D. Fuchs launched a line of research to address the issue of teachers’ nonresponse to CBM data. Specifically, they examined ways in which computers could be used to enhance teachers’ CBM implementation and use in reading, spelling, and mathematics. They examined the use of the software for automatically collecting, scoring, and graphing the data, and for providing diagnostic feedback and instructional recommendations to teachers (see L. Fuchs & Fuchs, 1989, 2002; L. Fuchs, Fuchs, Hosp, & Hamlett, 2003; Stecker et al., 2005).

Results from the initial set of studies in the research demonstrated that computers could be used to improve teachers’ implementation of data decision rules (i.e., changing instruction or raising the goal); however, teachers continued

to experience difficulty knowing specifically how to change, and what to change in, instruction (L. Fuchs & Fuchs, 2002; Stecker et al., 2005). Thus, in subsequent studies, teachers were provided with computerized skills analyses that displayed students' level performance on curricular subskills. The addition of skills analyses improved teachers' decisions regarding *what* to teach: Teachers designed more specific instructional modifications, and effected higher rates of student achievement (L. Fuchs & Fuchs, 2002). However, skills analyses did not enhance teachers' ability to change *how* they taught. Thus, a third feature was added—expert systems feedback—to provide teachers with “expert-based” suggestions about how to change instruction for individual students. Expert systems improved teachers' ability to design more diverse instructional programs, and, in reading and mathematics, positively influenced student achievement (see L. Fuchs & Fuchs, 2002; Stecker et al., 2005). In the final phase of research, these features were included in programs designed to be applied in general-education classrooms to monitor progress for larger groups of students (L. Fuchs & Fuchs, 2002; Stecker et al., 2005).

Since the 1980s and 1990s, little research has focused on the role of the teacher as decision-maker in CBM progress-monitoring. There are several reasons to return to a focus on the teacher. First—and probably most obvious—technology has changed dramatically in the last 20 years. The research by Fuchs, Fuchs and colleagues involved software platforms that are now outdated. Technology today includes internet, smart phones, tablets, and more sophisticated software capabilities. Research is needed to determine how such technology can be used to enhance teachers' use of progress-monitoring data (see Espin, Chung, Foegen, & Campbell, in press, for examples). Second, development of CBM measures and procedures has expanded since the 1980s/1990s to both younger and older students (see Wayman, Wallace, Wiley, Tichá, & Espin, 2007). It is unclear to what extent the skills analyses and expert systems developed for elementary-school children apply to younger or older children. More research is needed on use of such features with younger and older children.

Finally, and most relevant for the research presented in this special issue, it is not clear whether computers could (or should) usurp the role of the teacher. Even with computer supports, the teacher remains an essential part of the instructional decision-making process, as evidenced by research that has shown that actively involving teachers in the data-based decision-making process enhances the use of computer supports (L. Fuchs, Fuchs, & Hamlett, 1989; see also L. Fuchs & Fuchs, 1989, for a discussion of this issue). There is a need for more research on teachers' involvement in the data-based decision-making process. A first step in such research could be to address the question of *why* teachers have difficulty with data-based decision-making.

Teachers' Interpretation, Understanding, and Use of CBM Progress Data

There are several potential reasons for teachers' difficulties with data-based decision-making. One potential reason was addressed by the research of Fuchs and Fuchs—that is, that

teachers do not know what instruction to change, or how to change instruction. A second potential reason is that teachers do not believe, or do not *want* to believe, CBM data. For example, Foegen, Espin, Allinder, and Markell (2001) found that preservice teachers doubted the validity of CBM reading scores, especially with regard to the progress of individual students. A third potential reason, and the focus of the study presented here, is that teachers are unable to understand and interpret CBM progress graphs. There has been little (if any) research on teachers' ability to understand and interpret CBM progress graphs. Given what is known about the difficulties that people experience with graph interpretation (Friel et al., 2001), it seems worthwhile to examine teachers' abilities to understand and interpret CBM progress graphs. The challenge, however, is how to conduct such research.

In the exploratory study reported in this article, we describe the development of a method for examining teachers' understanding and interpretation of CBM graphs. Drawing upon the decision-making and graph-reading literature, we use a think-aloud procedure to study teachers' ability to describe CBM graphs, and examine whether the coherence, specificity, reflectivity, and accuracy of teachers' think-alouds are related to expert ratings of teachers' understanding of CBM.

METHOD

Participants

Participants in the study were 14 special education teachers (13 female) from an urban district. To recruit participants, district liaisons initially identified 52 special education teachers with experience using CBM. All 52 teachers were invited, but only 14 agreed to participate. Participants had between 1 and 31 years of teaching experience, and reported that they had between 1 and 20 years of experience using CBM, and that they had generated between 10 and 250 CBM graphs. Most teachers had learned how to implement CBM via district training. Teacher characteristics are presented in Table 1.

Think-Aloud Procedures

We used a think-aloud procedure to examine teachers' understanding and interpretation of CBM data. Each teacher was presented with three CBM graphs that displayed reading-aloud progress data from three students in the upper elementary grades (grades 3–5). The graphs were of actual student data, but not from the teachers' own students. The graphs displayed baseline data, a long-range goal, a goal line, and progress data across several instructional phases. Teachers were provided with pencils so they could draw in slope lines if they wished.

For the first two graphs, an open-ended think-aloud structure was used. Teachers were given the following prompt: *We want you to look at this graph of a special education student's progress in reading. Think out loud as you look at the graph and tell me what you are seeing and thinking. Tell me what you are looking at and why you are looking at it.* The order of the first two graphs was counterbalanced

TABLE 1
Teacher Characteristics

| Group | Weighted Mean Rating | Teaching License(s) | Highest Degree | Teaching Exp. (years) | Teaching Exp. in District (years) | CBM Exp. (years) | Number of CBM Graphs | CBM Training |
|--------|----------------------|-----------------------|----------------|-----------------------|-----------------------------------|------------------|----------------------|---------------------|
| High | 3.96 | LD/EBD/Social Studies | Master's | 32 | 31 | 20 | 250 | District |
| | 3.83 | LD/EBD/MMI/Elementary | Master's | 24 | 20 | 18 | 180 | University/District |
| | 3.46 | LD | Master's | 2 | 2 | 3 | 13 | University |
| | 3.37 | LD/EBD/Elementary | Master's | 27 | 27 | 20 | 200 | District |
| | 3.34 | LD/EBD/Elementary | Master's | 7 | 1 | 1 | 18 | District |
| Middle | 3.14 | LD/EBD/Elementary | Master's | 9 | 9 | 9 | 100 | District |
| | 3.04 | LD/Elementary | Master's | 14 | 14 | 14 | 280 | District |
| | 2.98 | LD/EBD | Bachelor's | 8 | 6 | 5 | 56 | District |
| | 2.96 | LD/EBD | Master's | 30 | 11 | 11 | 150 | District |
| Low | 2.38 | DCD | Master's | 1 | 1 | 1 | 10 | University/District |
| | 2.14 | LD/EBD/MMI | Master's | 21 | 17 | 14 | 240 | District |
| | 2.02 | LD/EBD/MMI/Elementary | Master's | 30 | 30 | 25 | 300 | District |
| | 1.85 | EBD/MMI | Master's | 14 | 1 | 10 | NR | University |
| | 1.82 | MMI/Elementary | ABD | 27 | 17 | 14 | 280 | District |

Note. Exp. = experience; NR = not reported; ABD = all but dissertation.

TABLE 2
Think-Aloud Prompts Used for Graph 3

| Area | Question(s) |
|-----------------|---|
| Administrator: | "We will now ask you a series of questions about this graph. We will ask you all the questions in a predetermined order. If you feel you have already answered a question, please elaborate on what you previously said." |
| Set up of Graph | <ul style="list-style-type: none"> What is this graph telling you about? |
| Baseline Data | <ul style="list-style-type: none"> How would you describe where the student starts/begins? How would you describe the student's initial progress? How would you describe the procedures used to establish the student's initial level of performance? Why? Do you think the data are adequate for making decisions about the student's initial level of performance? Why or why not? |
| Goal Line | <ul style="list-style-type: none"> Can you talk about the goal set for the student? What do you think about the expected rate of progress set for the student? Why? How would you describe the procedures used to set the goal for the student? |
| Interventions | <ul style="list-style-type: none"> How would you describe the process used for making a judgment about the effectiveness of the interventions? What impact if any did the interventions have on the student's level of performance? Why? What impact if any did the interventions have on the student's rate of progress? Why? What are your thoughts about the length of the interventions? Do you think the data are adequate for making decisions about the effectiveness of the interventions? Why or why not? |
| Goal Attainment | <ul style="list-style-type: none"> What are your thoughts about the student's progress compared to his/her goal? Why? Does it appear that the student will make the goal set by the teacher? Why or why not? Do you think the data are adequate for making decisions about the student's progress? Why or why not? |
| Data Analysis | <ul style="list-style-type: none"> Do you think the data are adequate for making decisions about the student's overall instructional program? Why or why not? |

across teachers. For the third graph, a directed think-aloud structure was used in which teachers were asked specific questions about different aspects of the progress-monitoring graph (see Table 2).¹ This graph was presented last to the teachers. The first two graphs displayed five and six instructional phases, respectively. The third graph displayed three

instructional phases. Data from the open-ended think-alouds (the first two graphs) were coded to examine teachers' understanding and interpretation of CBM graphed data. Data from the directed think-aloud (third graph) were used only as a source of information for the expert ratings, and were not coded.

TABLE 3
Items on the Teachers Interpretation Rating Scale (TIRS)

How Well Did the Teacher do Each of the Following:

1. Understand and interpret the CBM graphs overall?
2. Understand the meaning of the number of words read correctly in 1 minute?
3. Understand the setup of the CBM graphs?
4. Understand the reason for collecting baseline data?
5. Interpret baseline data?
6. Understand the function of the goal line?
7. Understand the procedures for constructing a goal line?
8. Accurately interpret the data to evaluate the effectiveness of interventions?
9. Accurately interpret the data to evaluate the appropriateness of changing the interventions?
10. Determine whether or not students attained their goals?
11. Understand and use level of performance in analyzing the data?
12. Understand and use slope/rate of performance in analyzing the data?
13. Understand and use variability in performance in analyzing the data?

Note. Experts rated teachers for each item on a scale of 1–4; 1 = not at all; 2 = somewhat; 3 = well; 4 = very well.

Expert Ratings of Teachers' Understanding and Interpretation of CBM Graphed Data

Verbatim transcriptions of the open-ended and directed think-alouds were given to four CBM experts who rated the think-alouds according to the extent to which they reflected understanding and interpretation of CBM graphed data. Expert raters were identified based on their work and research related to CBM progress monitoring. All experts had conducted and published research and conducted teacher training on CBM for more than 20 years, and were nationally recognized CBM experts. The ratings were blinded: Experts were not aware of the identity or characteristics of the teachers when conducting the ratings. For each teacher, raters read through the think-alouds (both open-ended and directed), and then rated the teacher's understanding and interpretation of the CBM graphs using the Teacher Interpretation Rating Scale (TIRS; see Table 3). Raters first provided a *global rating* (Item 1) and then rated understanding and interpretation of specific aspects of the graphs (Items 2–13). Ratings were on a four-point scale (4 = better understanding/interpretation; 1 = poorer understanding/interpretation). Raters were encouraged to provide comments to accompany their ratings. Each think-aloud transcript was rated by at least two experts. Mean scores across the two experts were used for analyses. For two randomly selected teachers, think-alouds were rated by all four experts to examine interrater agreement.

Think-Aloud Coding

Coding Categories

The initial focus of our coding was on the coherence and specificity of the open-ended think-alouds. Recall that experts attend to relevant information and build integrated and coherent knowledge structures (Ericsson, 2006a; Feltovich

et al., 2006); thus, we expected that higher-rated think-alouds would refer more often to relevant information and would reflect a more integrated and coherent understanding of CBM than lower-rated think-alouds. To examine the accuracy of this hypothesis, a sequential coherence coding system was developed.

Sequential coherence was defined as the extent to which think-alouds reflected a coherent, integrated, specific explanation of the progress-monitoring process. Development of the coding system was based on processes outlined in Ericsson and Simon (1993) and Ericsson (2006b) for coding think-aloud data, and involved several phases. In the first phase, two researchers with several years of experience in training teachers to use CBM independently created a list of the steps necessary to effectively create and use CBM graphed data for evaluation of student growth and instructional effectiveness. The researchers met and discussed the steps. There were few differences in the steps identified by each researcher. The few differences that did exist were discussed, and an initial set of coding categories was created. Coders then read the think-alouds, as well as the raters' comments to further revise and refine the coding system. The final coding system consisted of the following sequence of steps: (1) framing the data (FR); (2) describing baseline data (BL); (3) describing goal setting (GS); (4) describing data or progress in individual phases (P1-5 for Graph A; P1-6 for Graph B); and (5) evaluating goal attainment (GA). This sequence of steps was used as an ideal sequence of steps; that is, the sequence of steps a teacher would follow in setting up and interpreting data on a CBM graph to inform instruction (see Table 4 for definitions and examples of statements for each code).

Think-alouds were also coded for *specificity*. Teachers sometimes described progress in a general rather than a specific manner. An example of a general statement is, "The student showed good progress throughout the year." Specific statements, in contrast, referred to progress within a phase—for example, "The student made good progress with Intervention 1." A separate code was created for general progress (GP) statements, and the number of sequences involving the GP code was calculated and analyzed separately.

Two additional coding categories arose from initial reading of the think-alouds: *reflectivity* and *accuracy*. First, we observed that teachers often made observations about the information on the graph. We coded these statements as reflective statements (RS; see Table 4). Second, we observed that, across all coding categories, teachers' statements often were inaccurate—for example, stating that a student was progressing when the data showed no progress. Thus, after the statements were assigned a content code, they were coded as either accurate or inaccurate.

Coding Procedures

After the coding system had been developed, two coders worked together to parse each open-ended think-aloud into idea units. Idea units referred to a unit of text expressing one idea. After completing the parsing, the coders coded four think-alouds together. The remaining protocols were coded

TABLE 4
Coding for Sequential Coherence of Think-Alouds

| <i>Code</i> | <i>Definition</i> | <i>Examples</i> |
|---|---|--|
| FR (Framing the Data) | Teacher provides a framework for data interpretation. Includes reference to graph set up, measures, or scores. | <ul style="list-style-type: none"> Well it looks like I'm looking at a reading graph with the number of words read correctly in a minute. |
| BL (Baseline) | Teacher describes baseline data (beginning level of performance), including where or what data are, or process for establishing baseline. | <ul style="list-style-type: none"> This student's baseline was about 45 words in a minute. |
| GS (Goal Setting) | Teacher discusses long-range goal (LRG; expected ending level of performance) or short-term objective (STO; expected rates of growth). Includes where or what LRG/STO is, or process for setting LRG/STO. | <ul style="list-style-type: none"> They set their goal just under 80 words in a minute for the 32nd week. |
| P1–P5/P6 (Describe data or progress in individual phase.) | Teacher discusses growth, performance, progress, variability, etc., in Phase 1, 2, 3, 4, 5, or 6. | <ul style="list-style-type: none"> The fourth intervention really seemed to work, really going up. (P4) |
| GA (Goal Attainment) | Teacher discusses whether or not student has achieved the long-range goal. | <ul style="list-style-type: none"> The student did make progress, by the end of the year she was above her aim line. |
| GP (General Progress) | Teacher discusses data or progress in general without focusing on a particular phase. Coded when teacher refers to entire graph rather than a specific phase. | <ul style="list-style-type: none"> The student showed good progress throughout the year. |
| RS (Reflective Statement) | Teacher makes a reflective comment about information on graph. Includes comments on: actions/decisions of teacher, reasons for improvement/nonimprovement, material or procedures used for monitoring, or other comments based on personal experiences with students. | <ul style="list-style-type: none"> I don't know why they made that change. Two words a week growth is good for his age. |

separately by both coders. Intercoder agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements. Average agreement between the two coders was 85.28 percent (range 61.54 percent to 95.65 percent). Agreement for all but two protocols was above 80 percent. Disagreements on these two protocols were related primarily to differences in interpretation of the GP code. This definition was discussed and clarified, and all think-alouds were recoded. Any other disagreements were discussed until an agreement was reached. After coding the think-alouds for sequential coherence, specificity, and reflectivity, coders marked each statement as accurate or inaccurate.

Determining Sequential Coherence

Sequential coherence was determined by calculating the percent of statement sequences following the ideal sequence set by the researchers. A sample of the form used for determining sequential coherence is presented in Figure 2. Along the top row and down the left column of the figure are the codes for each step in the ideal sequence. The last code is the GP code, which is shaded to reflect the fact that a sequence involving a GP statement was recorded but was treated separately in the analysis. Note that reflective statements were not included in the sequential coherence coding. These statements were counted separately.

The sequential coding proceeded from left to right; thus, if the teacher made a statement about framing the data, and this statement was followed by a statement about baseline scores, a tally mark was made at the intersection of FR (left) and BL (top). If the teacher then followed the baseline com-

ment with a comment about goal setting, a tally mark was made at the intersection of BL (left) and GS (top). Once the sequential coherence coding form was completed, the percentage of sequential statements at, above, and below the diagonal was counted.² Sequential statements *at*—that is, directly above—the diagonal (lightly shaded in figure) were those that followed the ideal sequence: The higher the number, the greater the sequential coherence of the think-aloud. To illustrate the coding process, an example of a coded think-aloud is presented in the Appendix. (The sequential coding for that think-aloud is presented in Figure 2.)

Analyses

The number of sequential statements at, above, and below the diagonal and the number of sequences with a GP code were compared for teachers with higher and lower expert ratings. In addition, the number of reflective statements and the number of accurate/inaccurate statements were compared. Because statements were within teacher, we used an analysis that accounted for dependency in the data. Specifically, we used a test for the association term in a Poisson generalized linear model (Agresti, 2007), which is equivalent to Pearson's Chi-square. This generalized linear model was adapted for correlated responses in generalized estimating equations (GEEs; Liang & Zeger, 1986). We used these GEEs where we report the test of the association parameter. This is a Wald type of Chi-square statistic, which we will denote by $W_{(p)}$, where p refers to the number of degrees of freedom. Analyses were performed in SPSS version 18, in which we used a robust covariance estimator and an exchangeable working correlation.

| | <i>FR</i> | <i>BL</i> | <i>GS</i> | <i>P1</i> | <i>P2</i> | <i>P3</i> | <i>P4</i> | <i>P5</i> | <i>GA</i> | <i>GP</i> |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <i>FR</i> | | | | | | | | | | |
| <i>BL</i> | | | 1 | | | | | | | |
| <i>GS</i> | | | | 1 | | | | | | |
| <i>P1</i> | | | | | 1 | | | | | |
| <i>P2</i> | | | | | | 1 | | | | |
| <i>P3</i> | | | | | | | 1 | | | |
| <i>P4</i> | | | | | | | | 1 | | |
| <i>P5</i> | | | | | | | | | 1 | |
| <i>GA</i> | | | | | | | | | | |
| <i>GP</i> | | | | | | | | | | |

FIGURE 2 Form for coding sequential coherence.
 Note. FR = framing the data; BL = baseline; GS = goal setting; P1–P5 = Phase 1 through Phase 5; GA = goal attainment; GP = general progress.

RESULTS

Expert Rater Groupings

Experts tended to agree on their teacher ratings. Global ratings (Item 1) were identical or within 0.5 for each expert pair, with the exception of two ratings where the global ratings were 3 and 4 and 2 and 4, respectively. For the pair with global ratings of 2 and 4, ratings on all other items were identical or within one point of each other. For the two teachers whose think-alouds were rated by all four experts, the percent of items for which three of the four experts assigned a rating of either 1/2 or 3/4 was calculated. Percent agreement was 91 percent. Note that it was not always the same expert who was in disagreement with the other experts.

Using the expert ratings, teachers were divided into high, middle, and low groups on the basis of a weighted mean score. The weighted mean was calculated by adding the global rating (Item 1) to the mean rating for the remaining items (Items 2–13) and dividing by two. By taking this approach, equal weight was given to the global rating and mean ratings across individual items. The weighted means are reported in Table 1 with the teacher characteristics. The global rating, mean ratings across items 2–13, and ratings on each item of the *TIRS* are displayed in Table 5 for teachers in the high, middle, and low groups (ordered according to the weighted means). The number and short description at the top of the table refer to each item rated by the experts (see also Table 3). The last column presents the mean rating for items 2–13.

Inspection of Table 1 reveals little correspondence between ratings and the descriptive characteristics of the

teachers, including self-reported experience with CBM use. Correlations between the weighted means and teacher characteristics were: years of teaching experience ($r = -.02$), years of teaching experience within the district ($r = .16$), years of experience using CBM ($r = -.01$), and number of CBM graphs created ($r = -.29$). All correlations were nonsignificant.

Inspection of Table 5 reveals a high degree of correspondence between the global ratings (Item 1, column 1) and mean ratings of items 2–13 (last column); not surprisingly, the correlation between the global rating (Item 1) and mean ratings (Items 2–13) was large ($r = .98, p < .001$). The means for each group are displayed in bold in Table 5. Items assigned a rating lower than 3 for individual teachers are shaded in gray. Examination of the group mean scores across each item reveals a pattern of clear and consistent differences between the high- and low-rated groups. Mean ratings for the high-rated group are all above 3, while mean ratings for the low-rated group are all below 2.5, with a difference of at least one rating point separating the two groups on every item. Mean ratings for the middle-rated group are more variable, with some similar to the high-rated group and others similar to the low-rated group. Item 12 (understanding/use of slope) was rated low across all three groups, and Item 10 (goal attainment) was rated high across all three groups.

Think-Alouds: Coherence, Specificity, Accuracy, and Reflectivity

To examine *sequential coherence*, the number and percentage of statement sequences that fell at, above, and below

TABLE 5
Item Ratings for Individual Teachers and Groups

| Group | Global Rating (1) | Meaning WRC (2) | Baseline | | Goal Line, Function | | Goal Line, Construct. | | Interpret Data, Effect | | Interpret Data, Change Interv. (9) | Goal Attainment (10) | Under./Use, Level (11) | Under./Use, Slope (12) | Under./Use, Variab. (13) | Mean Items 2-13 |
|-------------|-------------------|-----------------|----------------|----------------|---------------------|----------------|-----------------------|-------------|------------------------|-------------|------------------------------------|----------------------|------------------------|------------------------|--------------------------|-----------------|
| | | | Reason For (4) | Interpret. (5) | Function (6) | Construct. (7) | Interpret. (8) | Effect (8) | | | | | | | | |
| High | 4 | 3.5 | 4 | 3.5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3.92 |
| | 4 | 4 | 3.5 | 3.25 | 3.75 | 4 | 4 | 4 | 3.5 | 3.5 | 3.5 | 4 | 3.5 | 4 | 3 | 3.65 |
| | 3.5 | 3 | 4 | 3.5 | 3.5 | 3.5 | 3.5 | 4 | 3 | 3.5 | 3.5 | 4 | 3.5 | 2.5 | 3 | 3.42 |
| | 3.25 | 4 | 3.5 | 3.5 | 4 | 3.5 | 3.5 | 4 | 3.75 | 3.5 | 3.5 | 4 | 3 | 2 | 3.5 | 3.48 |
| | 3.25 | 3.5 | 3.5 | 3.5 | 3.5 | 4 | 3.5 | 4 | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 | 2.5 | 3 | 3.42 |
| Mean | 3.6 | 3.6 | 3.7 | 3.45 | 3.75 | 3.8 | 3.8 | 3.55 | 3.6 | 3.6 | 3.9 | 3.5 | 3.5 | 3.0 | 3.3 | 3.58 |
| Middle | 3 | 1.5 | 4 | 3.5 | 4 | 4 | 4 | 3.5 | 3.5 | 2.5 | 4 | 4 | 3 | 2.5 | 4 | 3.27 |
| | 3 | 3 | 3.5 | 2.25 | 4 | 3.5 | 3.5 | 3 | 2.5 | 3 | 3.5 | 3.5 | 3 | 2 | 3 | 3.07 |
| | 3 | 2 | 4 | 2.5 | 2.25 | 2.25 | 4 | 3.25 | 3.25 | 2.25 | 4 | 4 | 3.5 | 3 | 2 | 2.95 |
| | 3 | 2.5 | 3 | 3 | 3.5 | 3.5 | 3 | 3 | 3 | 2.5 | 3 | 3 | 2.5 | 2.5 | 3 | 2.92 |
| | Mean | 3 | 2.25 | 3.38 | 2.81 | 3.5 | 3.31 | 3.06 | 3.06 | 2.56 | 3.63 | 3.63 | 3.0 | 2.5 | 3.0 | 3.05 |
| Low | 2 | 2 | 2 | 1.5 | 4 | 3.5 | 3 | 3 | 3 | 3 | 3 | 3 | 3.5 | 2 | 3.5 | 2.76 |
| | 2.25 | 2 | 2 | 1.5 | 2 | 2 | 1 | 2.5 | 2.25 | 2.25 | 3.5 | 3.5 | 3 | 1 | 2.5 | 2.02 |
| | 2 | 2 | 2.5 | 2 | 2 | 2 | 1.5 | 2 | 2 | 2 | 2.5 | 2 | 2 | 2 | 2 | 2.04 |
| | 2 | 2 | 1.5 | 1.75 | 1.5 | 1.25 | 1.5 | 1.5 | 2 | 2 | 3 | 3 | 1.5 | 1.5 | 1 | 1.70 |
| | 1.75 | 1.5 | 2 | 1.5 | 2.5 | 2 | 2 | 2 | 1.5 | 1.5 | 2 | 2 | 2 | 1.5 | 2 | 1.88 |
| Mean | 2 | 1.9 | 2.1 | 1.55 | 2.4 | 1.85 | 2.2 | 2.2 | 2.15 | 2.8 | 2.8 | 2.4 | 1.6 | 2.2 | 2.08 | |

Note. Numbers in parentheses are item numbers. WRC = words read correctly; Interpret. = interpretation; Construct. = construction; Interven. = intervention; Under. = understanding; Variab. = variability. Means displayed in bold. Ratings below 3 shaded.

TABLE 6
Number and Percentage of Sequences Below/On, At, or Above Diagonal

| Group | Below/On Diagonal | At Diagonal | Above Diagonal | Total Sequences |
|--------|-------------------|-------------|----------------|-----------------|
| High | 2 (3.23) | 55 (88.71) | 5 (8.06) | 62 (100) |
| Middle | 11 (25.58) | 24 (55.81) | 8 (18.60) | 43 (100) |
| Low | 4 (14.81) | 15 (55.56) | 8 (29.63) | 27 (100) |

Note. Percentages are in parentheses. Sequences with reflective statements or general progress statements not included in the total.

TABLE 7
Number and Percentage of General Progress Sequences, Accurate Statements, and Reflective Statements

| Group | General Progress Sequences ^a | Accurate Statements ^b | Reflective Statements ^b |
|--------|---|----------------------------------|------------------------------------|
| High | 10 (13.89) | 114 (98.28) | 34 (29.31) |
| Middle | 13 (23.21) | 76 (97.44) | 14 (17.95) |
| Low | 23 (46.00) | 56 (60.22) | 33 (35.48) |

Note. Percentages are in parentheses.

^aSequences that included general progress statements. Total number of sequences were 72, 56, and 50 for high, middle, and low groups, respectively. (Sequences with reflective statements not included in total.)

^bTotal number of statements were 116, 78, and 93 for high, middle, and low groups, respectively.

the diagonal for each group were compared for high, middle, and low groups (see Table 6). The greater the number of sequences at the diagonal, the greater the sequential coherence of the think-aloud. Chi-square analysis revealed significant group differences on sequential coherence ($W_{(4)} = 17.63, p = .001$). The high-rated group had the largest percentage of statement sequences at the diagonal (nearly 89 percent), while the middle- and low-rated groups had a much lower percentage of sequences at the diagonal (56 percent each).

To examine *specificity*, the number and percentage of statement sequences that included a general progress code (GP) were compared (see Table 7). Chi-square analyses revealed significant group differences ($W_{(2)} = 9.50, p = .009$), with teachers in the low-rated group having a substantially higher percentage of statement sequences with GP codes (46 percent) than teachers in the middle- (23 percent) or high-rated (14 percent) groups.

To examine the *reflectivity* and *accuracy* of the think-alouds, we focused on the statements themselves rather than the statement sequences. Significant, and striking, group differences were seen in the number and percentage of accurate statements ($W_{(2)} = 42.87, p = .001$ see Table 7), with only 60 percent of the statements made by teachers in the low-rated group coded as accurate, compared to 97–98 percent for teachers in the middle- and high-rated groups. There were also significant group differences in the percentage of reflective statements: Teachers in the high- and low-rated groups made reflective statements 29 and 35 percent of the time, respectively, compared to 18 percent for the middle-rated group ($W_{(2)} = 12.07, p = .002$; see Table 7). However, examination of the accuracy of the reflective statements reveals that 42 percent of the reflective statements made by the low-rated group were not accurate, compared to only 6 percent for the high-rated group and 0 percent for the middle-rated group.

DISCUSSION

Our findings suggest that a think-aloud methodology can be used to examine teachers' understanding and interpretation of CBM progress graphs, and that coding think-alouds for coherence, specificity, accuracy, and reflectivity may serve to differentiate teachers with more and less skill in understanding/interpreting the graphs. The think-alouds for teachers with higher ratings from the CBM experts were more coherent, specific, and accurate than were the think-alouds for teachers with lower ratings. In addition, the higher-rated teachers made a greater number of accurate reflective statements. Think-alouds for teachers in the middle-rated group were less coherent than, but as specific and accurate as, think-alouds for the high-rated group.

Teachers who were rated by experts as high in CBM knowledge seemed quite competent in describing and interpreting CBM progress data. These teachers talked about the data in a coherent, specific, and accurate manner, making accurate, reflective statements about the graph content and the processes they assumed were used to collect the data. These teachers seemed to be at the level of *reading beyond the data* as described by Curcio (1987) and Friel et al. (2001); that is, they were able to select relevant information from the graphs, and to integrate and interpret relations in the data to draw inferences to solve problems or answer questions.

In contrast, teachers who were rated by experts as low in CBM knowledge seemed to lack competence in interpreting and describing CBM progress data. These teachers did not talk about the data in a coherent, specific, or accurate manner; and when they reflected upon the data or the processes used to collect the data, their reflections were often inaccurate. The inaccuracy of many of their statements raises the question as to whether these teachers had even reached the beginning level of graph interpretation, *reading the data* (Curcio, 1987;

Friel et al., 2001). For some of these teachers, extracting and describing the data accurately was difficult.

Teachers in the middle-rated group described CBM progress data in a less coherent manner than did the high-rated group, but described the data in a specific and accurate manner and, when they reflected upon the data or processes used to collect the data, did so accurately. These teachers demonstrated the ability to *read the data*, and, to a certain extent, to *read between the data* (Curcio, 1987; Friel et al., 2001)—that is, to interpret and find relations in the data, even though they did not reflect upon the data as often as the higher-rated teachers. Of course, one must keep in mind that these observations are drawn from data from a small sample and are meant only to provide a framework for future research.

The lack of correspondence between teachers' self-reported experience using CBM and experts' ratings of understanding and interpretation was surprising. Perhaps teachers reported using CBM for ongoing progress monitoring when they actually used it primarily for district-wide screening or benchmarking. Arguing against this explanation, however, is the low and nonsignificant correlation between expert ratings and number of graphs that teachers reported creating. Three of the five teachers with the lowest ratings reported creating between 240 and 300 CBM graphs over the course of their careers. If replicated, these data would suggest that experience with CBM does not guarantee that teachers can adequately interpret and describe CBM graphed data, at least not when asked to do so via a think-aloud format with researcher-provided graphs. Future research should examine the relation between teachers' ability to interpret and describe CBM data that have been collected from their own students.

Although used as a method to separate the think-alouds in our study, two observations are in order regarding the expert ratings. First, on the item, "Understand and use slope/rate of improvement," nearly all teachers received ratings of lower than 3. Second, on the item, "Determine whether or not students attained their goals," nearly all teachers received ratings of 3 or higher. It is possible that ratings on these items were influenced by the graph selection. For example, with regard to slopes, it is possible that teachers thought they should not consider or refer to slopes because the slopes were not drawn on the graphs. With regard to goal attainment, it was fairly easy in all three graphs to determine whether or not students had achieved their goal. However, if results related to these two items were to replicate, it would be "bad news" for CBM progress monitoring. Understanding and using slope is essential if teachers are to use progress data to evaluate the effectiveness of instructional programs, and to change those programs when necessary. While it is encouraging to see that even the lowest rated teachers were able to determine whether or not students had attained their long-range goal, this component of CBM data analysis is one of the least important aspects of using progress monitoring for building effective instructional programs. Decisions related to goal attainment take place at the end of the academic year, when it is too late to change ineffective instructional programs. In future research, it would be worthwhile to examine directly the interpretation of various slope and goal patterns.

The findings of this exploratory study align well with the descriptions of expertise by Ericsson (2006a) and Feltovich et al. (2006). First, *experience and expertise are not the same*. Our data revealed a lack of correspondence between self-reported experience with CBM and ratings of understanding and interpretation of CBM graphed data. Second, *knowledge of content matter (in this case, CBM) is important to expertise*. Teachers who received the highest ratings in our study made few if any inaccurate statements in their think-alouds, and made accurate reflections about the processes they assumed were used in the collection and use of the data.

Third and fourth, *expertise involves selective access to important and relevant information, and experts not only know more, they know different—they build integrated and coherent knowledge structures of the field*. In our study, high-rated teachers produced specific, integrated, and coherent think-alouds. Nearly all statement sequences for high-rated teachers followed an "ideal" sequence for data-based instructional decision-making, and their descriptions of student progress were specific to individual instructional phases. Teachers in the middle-rated group produced less coherent think-alouds, but also tended to be specific in their descriptions. Teachers in the low-rated group produced think-alouds that were neither coherent nor specific. Given the consistency of our results with both the graph-interpretation and expertise literature, we believe that we may be on the right path in our desire to study teachers' understanding and interpretation of graphed CBM data, and in our approach for doing so.

LIMITATIONS

A major limitation to the study was the small sample size and likely selection bias. Only 14 of 52 teachers agreed to participate in the study. It is possible that teachers who agreed to participate were those who were positively disposed to CBM, and who viewed themselves as having a fairly good grasp of CBM. If so, then our results might reflect a higher level of understanding and interpretation than would be found in a representative sample of teachers. Given the small, potentially biased sample, the findings of this study are best viewed as a starting point for future research. Nonetheless, the study does make two important contributions: (1) the development of a methodology for the study and analysis of teachers' understanding of graphed data, and (2) highlighting the need for research on teachers' understanding and interpretation of graphed data by demonstrating that teachers can have great difficulty with these skills.

A second limitation is that the graphs used for the think-alouds included instructional changes made at preset times, as opposed to changes made in response to student progress data. We used these graphs because they were graphs of real student data and included several instructional changes throughout the school year. In addition, they provided an opportunity for teachers to reflect on the appropriateness of instructional changes. On the other hand, the timing of the instructional changes may have confused teachers during think-aloud sessions. In future research, it would be worthwhile to create a set of graphs specifically designed

to study teachers' understanding and interpretation of CBM data.

A third limitation is that teachers did not examine graphs of their own students' progress. It is possible that a different pattern of results would emerge if teachers were to talk about data from their own students. In future research, it will be important to study teachers' understanding and interpretation of progress data collected from their own students, and, ultimately, to examine whether differences in understanding and interpretation of such graphed data relate to differences in teachers' use of the data for instructional decision-making.

IMPLICATIONS AND FUTURE DIRECTIONS FOR RESEARCH

Results of this exploratory study suggest that researchers can use a think-aloud approach by asking participants to describe graphed data and then analyze think-aloud transcripts for coherence, specificity, accuracy, and reflectivity, to study understanding and interpretation of CBM graphed data. This method might be used in the future to study the link between graph interpretation and the use of the data for decision-making, to identify teachers who are likely to experience difficulty with graph interpretation, to follow teachers' development of graph interpretation skills, to identify areas of graph interpretation in which additional training and practice may be needed by teachers, and, most importantly, to guide the development of methods for improving teachers' ability to understand and interpret CBM graphed data.

An important topic to be addressed in future research, and one that was raised in the introduction, is how new technologies can be used to support teachers' understanding, interpretation, and use of CBM data. We provide two examples of such research here.

Technology could be used to highlight relevant graph features, potentially aiding in graph interpretation. Research on graph comprehension shows that graph design affects the viewer's ability to understand and interpret graph content, and that using designs that highlight the important features of the graphs improves understanding and interpretation (see Fausset, Rogers, & Fisk, 2008; Friel et al., 2001). CBM graphing programs could highlight important graph features by providing a function that allows teachers to hide data points to display slope and goal lines only, or to extend the slope line to compare the actual rate of growth (slope) to the expected rate of growth (goal line).

Technology could also be used to strengthen the link between the CBM data and available instructional information. There are a multitude of sources available on the internet that provide information on evidence-based interventions—for example, the University of Missouri's Evidence-Based Network (<http://ebi.missouri.edu>) and What Works Clearinghouse (<http://ies.ed.gov/ncee/Wwc/>). The challenge that teachers face is efficiently and effectively accessing the right information at the right time for an individual student. Building upon the early work by Fuchs and Fuchs (see L. Fuchs & Fuchs, 1989, 2002; Stecker et al., 2005), technology could be designed to help guide teachers to information that is relevant for individual students.

CONCLUSION

Our results suggest that think-alouds can be used to study teachers' understanding and interpretation of CBM graphs, and that, for some teachers, this understanding and interpretation of CBM graphed data may be quite difficult. Replication of these results in future research would imply that we as researchers and teacher-educators have much work to do in the way of helping teachers to improve understanding and interpretation of CBM progress data, especially if we expect teachers to use those data to guide instructional decision-making for their students.

ACKNOWLEDGMENTS

Research supported by Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

NOTES

1. One teacher mistakenly received Graph A for the directed think-aloud. Because the primary focus of the analysis was the open-ended think-alouds rather than the directed think-alouds, the data for this teacher were maintained; however, it is possible that seeing the graph a second time influenced the teacher's responses to the directed think-aloud. In Table 7, it can be seen that this teacher (the second teacher listed in the low group) had a number of items that were rated as 3 or 4. This was the only teacher in the low group with numerous 3- and 4- ratings.
2. Because of the method used for parsing and coding, idea units rarely if ever fell on the diagonal. For example, a Phase 1 statement rarely followed another Phase 1 statement. If a teacher made two statements about Phase 1, the two statements most often reflected just one idea, and thus were coded as a single idea unit. However, if the two statements reflected two ideas, a tally was placed directly on the diagonal. In our data, this occurred only once: A Goal Statement was followed by a Goal Statement. We included this sequence in the below-the-diagonal totals.

REFERENCES

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley & Sons. doi:10.1002/0470114754.
- Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions*. London: Harper.
- Baron, J. (2000). *Thinking and deciding* (3rd ed.). New York: Cambridge University Press.
- Berliner, D. C. (1989). Implications of studies of expertise in pedagogy for teacher education and evaluation. In *New directions for teacher assessment: Proceedings of the 1988 ETS Invitational Conference* (pp. 39–67). Princeton, NJ: Educational Testing Service.

- Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. In J. H. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students* (pp. 161–186). Fort Worth, TX: Holt, Rinehart, & Winston.
- Berry Kuchle, L., Zumeta Edmonds, R., Danielson, L. C., Peterson, A., & Riley-Tillman, T. C. (2015). The next big idea: A framework for integrated academic and behavioral intensive intervention. *Learning Disabilities Research & Practice, 30*, 150–158. doi:10.1111/ldrp.12084
- Black, P., & Wiliam, D. (2005). Changing teaching through formative assessment: Research and practice. In *Formative assessment: Improving learning in secondary classrooms* (pp. 223–240). Centre for Educational Research and Innovation (CERI). Paris: Organisation for Economic Co-operation and Development.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155–166. doi:10.1016/j.learninstruc.2009.02.014
- Codding, R. S., Skowron, J., & Pace, G. M. (2005). Back to basics: Training teachers to interpret curriculum-based measurement data and create observable and measurable objectives. *Behavioral Interventions, 20*, 165–176. doi:10.1002/bin.194
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382–393. doi:10.2307/749086
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change, 17*, 7–28. doi:10.1007/s10833-015-9264-2
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Remedial and Special Education, 37*, 184–192. doi:10.1177/00224669030370030801
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.
- DeStefano, L., Shriner, J. G., & Lloyd, C. A. (2001). Teacher decision making in participation of students with disabilities in large-scale assessment. *Exceptional Children, 68*, 7–22. doi:10.1177/001440290106800101
- Ericsson, K. A. (2006a). An introduction to The Cambridge Handbook of Expertise and Expert Performance: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. J. Feltoovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 3–20). New York: Cambridge University Press. doi:10.1017/CBO9780511816796.001
- Ericsson, K. A. (2006b). Protocol analysis and expert thought: Concurrent verbalization of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltoovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223–242). New York: Cambridge University Press. doi:10.1017/CBO9780511816796.013
- Ericsson, K. A., Charness, N., Feltoovich, P. J., & Hoffman, R. R. (Eds.) (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press. doi:10.1017/CBO9780511816796
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Ericsson, K. A., & Ward, P. (2007). Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science, 16*, 346–355. doi:10.1111/j.1467-8721.2007.00533.x
- Espin, C. A., Chung, S., Foegen, A., & Campbell, H. (in press). Curriculum-based measurement for secondary-school students. In M. Kennedy & P. Pullen (Eds.), *RTI handbook*. New York: Routledge.
- Fausset, C. B., Rogers, W. A., & Fisk A. D. (2008). Visual graph display guidelines. Technical Report HFA-TR-0803. Atlanta: Georgia Institute of Technology, School of Psychology – Human Factors and Aging Laboratory. Retrieved November 1, 2016 from https://smartech.gatech.edu/bitstream/handle/1853/40573/HFA-TR-0803%20Graph_Guidelines_Final.pdf
- Feltoovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltoovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41–68). New York: Cambridge University Press. doi:10.1017/CBO9780511816796.004
- Foegen, A., Espin, C. A., Allinder, R. M., & Markell, M. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *Journal of Special Education, 34*, 226–236. doi:10.1177/002246690103400405
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124–158. doi:10.2307/749671
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449–460. doi:10.3102/00028312021002449
- Fuchs, L. S., & Fuchs, D. (1989). Enhancing curriculum-based measurement through computer applications: Review of research and practice. *School Psychology Review, 18*, 317–327.
- Fuchs, L. S., & Fuchs, D. (2002). Computer applications to curriculum-based measurement. *Special Services in the Schools, 17*, 1–14. doi:10.1300/1008v17n01_01
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Computers and curriculum-based measurement: Effects of teacher feedback systems. *School Psychology Review, 18*, 112–125.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Hamlett, C. L. (2003). The potential for diagnostic analysis within curriculum-based measurement. *Assessment for Effective Intervention, 28*, 13–22. doi:10.1177/073724770302800303
- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-Intervention: A decade later. *Journal of Learning Disabilities, 45*, 195–203. doi:10.1177/0022219412442150
- Garcia-Retamero, R., Okan, Y., & Cokely, E. T. (2012). Using visual aids to improve communication of risks about health: A review. *The Scientific World Journal, 2012*, 1–10. doi:10.1100/2012/562637
- Gigerenzer, G. (2007). Fast and frugal heuristics: The tools of bounded rationality. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment & decision making* (pp. 62–88). Malden, MA: Blackwell. doi:10.1002/9780470752937.ch4
- Koehler, D. J., & Harvey, N. (Eds.). (2007). *Blackwell handbook of judgment & decision making*. Malden, MA: Blackwell. doi:10.1002/9780470752937
- Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Statistical and methodological advances* (pp. 91–125). Washington, DC: American Psychological Association. doi:10.1037/14376-004
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22. doi:10.2307/2336267
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*, 71–85. doi:10.1080/00461520.2012.667064
- Mandinach, E. B., & Gummer, E. S. (2013). A systematic view of implementing data literacy in educator preparation. *Educational Researcher, 42*, 30–37. doi:10.3102/0013189X12459803
- Perkins, D. N. (2009). Decision making and its development. In E. Callan, T. Grotzer, J. Kagan, R. E. Nisbett, D. N. Perkins, & L. S. Shulman (Eds.), *Education and a civil society: Teaching evidence-based decision making* (pp. 1–28). Cambridge, MA: American Academy of Arts and Sciences.
- Sabers, D., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 28*, 63–88. doi:10.2307/1162879
- Schildkamp, K., Ehren, M., & Lai, M. K. (2012). Editorial article for the special issue on data-based decision making around the world: From policy to practice to results. *School Effectiveness and School Improvement, 23*, 123–131. doi:10.1080/09243453.2100.652122
- Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science, 3*, 560–578. doi:10.1111/j.1756-8765.2009.01066.x

- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19. doi:10.1146/annurev.ps.41.020190.000245
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*, 795–819. doi:10.1002/pits.20113
- Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, *24*(6), 9–17. doi:10.3102/0013189X024006009
- Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, *27*, 533–556. doi:10.3102/00028312027003533
- Tsui, A. B. (2003). *Understanding expertise in teaching: Case studies of second language teachers*. New York: Cambridge University Press. doi:10.1017/CBO9781139524698
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, *18*, 137–146. doi:10.1111/1540-5826.00070
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*, 85–120.
- Wickens, C. S., Lee, J., Liu, Y., & Gordon-Becker, S. (2004). *An introduction to human factors engineering* (2nd ed.). Upper Saddle River, NJ: Pearson /Prentice Hall.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Educational Policy Analysis Archives*, *18* (19), 1–37. doi:10.14507/epaa.v18n19.2010
- Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, *27*, 1073–1079. doi:10.3758/BF03201236

Appendix

TABLE A1
Sample Think-Aloud with Codes, Graph 1

| Teacher Statement | Code |
|---|--------------------------------|
| Okay. Now the student's baseline was about 45 words in a minute. | BL (Baseline Statement) |
| The goal was set to just under 80 words in a minute for the 32nd week. | GS (Goal Statement) |
| With intervention one, the student went up and down. | P1 (Progress, Phase 1) |
| Student did better with intervention two. There are more scores above the line than below the line. | P2 (Progress, Phase 2) |
| Intervention three was similar to intervention two, | P3 (Progress, Phase 3) |
| but the student really did well with intervention four. All of the scores are at or above the line. | P4 (Progress, Phase 4) |
| Intervention five did not work well. The student went back down below the line, | P5 (Progress, Phase 5) |
| and ended up reading below the goal, at about 65 words or so a minute I think. | GA (Goal Attainment Statement) |
| So if the teacher had had more time, you know, it was obviously the end of the year, but if it had been me, | RS (Reflective Statement) |
| I would have gone back to intervention four, because the student did well in intervention four. | |

Note. Sample think-aloud created on the basis of think-alouds from teachers in the high-rated group.

About the Authors

Christine A. Espin, Ph.D., is a Professor in the Department of Education and Child Studies (Special Education / Learning Disabilities) at Leiden University, The Netherlands. Her research interests include teachers' data-based decision-making, Curriculum-based Measurement in reading, writing, and content-area learning for secondary-school students with learning disabilities, and reading comprehension interventions for students with reading difficulties.

Miya Miura Wayman, Ph.D., earned her Ph.D. in Educational Psychology/Special Education from the University of Minnesota. Her current research interests include developing progress monitoring measures for beginning readers and improving teachers' use of data for instructional decision-making.

Stanley L. Deno, Ph.D., passed away in October of 2016. He was a Professor Emeritus of Educational Psychology/Special Education at the University of Minnesota. His career was dedicated to finding ways to help educators assess the academic progress of students with learning difficulties and use the progress data to build effective instructional programs for the students. He (in collaboration with his doctoral student, Phyllis K. Mirkin) was responsible for developing the idea that eventually became known as Curriculum-Based Measurement.

Kristen L. McMaster, Ph.D., is a Professor of Special Education in the Department of Educational Psychology, University of Minnesota. Her research interests include creating conditions for successful response to intervention of students at risk and students with disabilities. Specific research focuses on (1) promoting teachers' use of data-based decision making and evidence-based instruction and (2) developing intensive, individualized interventions for students for whom generally effective instruction is not sufficient.

Mark de Rooij, Ph. D., is Professor of Methodology and Statistics of Psychological Research in the Institute of Psychology at Leiden University. His research interest are in statistical modelling of longitudinal data, in statistical learning, and in the distinction between prediction and explanation.