7-1-2015

Can Teachers Accurately Predict Student Performance?

Keith W. Thiede
Boise State University

Jonathan L. Brendefur
Boise State University

Richard D. Osguthorpe
Boise State University

Michele B. Carney
Boise State University

Amanda Bremner
Boise State University

See next page for additional authors

Publication Information

Thiede, Keith W.; Brendefur, Jonathan L.; Osguthorpe, Richard D.; Carney, Michele B.; Bremner, Amanda; Strother, Sam; Oswalt, Steven; and Snow, Jennifer L.. (2015). "Can Teachers Accurately Predict Student Performance?". *Teaching and Teacher Education, 49*, 36-44. http://dx.doi.org/10.1016/j.tate.2015.01.012

Authors
Keith W. Thiede, Jonathan L. Brendefur, Richard D. Osguthorpe, Michele B. Carney, Amanda Bremner, Sam Strother, Steven Oswalt, and Jennifer L. Snow

Can teachers accurately predict student performance?

Keith W. Thiede

Jonathan L. Brendefur

Richard D. Osguthorpe

Michele B. Carney

Amanda Bremner

Sam Strother

Steven Oswalt

Jennifer L. Snow

Boise State University

John Sutton

Dan Jesse

RMC Research Corporation

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., Oswalt. S., Snow, J.L., Sutton, J., & Jesse, D. (2015). Can teachers accurately monitor student learning? *Teaching and Teacher Education*, 49, 36-44.

Abstract

In two studies, we examined the effect of professional development to improve mathematics instruction on the accuracy of teachers' monitoring of student learning. Study 1 was conducted with 36 teachers participating in three years of professional development. Judgment accuracy was influenced by the fidelity with which what was learned in the professional development. Study 2 was conducted with 64 teachers from 8 schools, which were randomly assigned to receive professional development or serve as a control. Judgment accuracy was greater for teachers receiving professional development than for teachers who did not and teachers were better to predict students' computational skills.

Keywords: teacher judgment; judgment accuracy; mathematics achievement

1. Introduction

Teachers can engage in many metacognitive and cognitive activities that are relevant to guiding students' learning (for a review, see Duffy, Miller, Parsons, & Meloth, 2009). For instance, a teacher may notice a particular student is struggling to finish an in-class activity, and the teacher may then attempt to implement strategies to help the student perform better. There is no doubt that effective teachers make many decisions each day (some estimates are in the thousands, e.g., Doyle, 1977) about how to restructure their pedagogy to meet the needs of individual students (Donovan, Bransford, & Pellegrino, 1999). Effectively adapting one's teaching to individual learners necessitates teachers accurately evaluating students' on-going states of mind or performance (Shavelson, 1978). The question here is straightforward: How well can a teacher evaluate the degree to which students have learned the target materials? For instance, can teachers accurately predict how well each student will perform on an upcoming exam? Certainly such a prediction would comprise just a small portion of all the activities a teacher could engage in to evaluate student learning, but if teachers can make these predictions accurately, it would be an invaluable skill to help them better regulate their students' learning. In this era of increased accountability, witnessed in several countries, teachers are expected to make more accurate evaluations about their students' understanding, and drawing on their diagnosis, better support their students during instruction; thus, finding ways to improve the accuracy of teachers' judgments of learning is particularly important.

In the remainder of this introduction, we will describe how researchers have measured the accuracy of teachers' judgments and discuss some of the pitfalls of interpreting these data. We will then provide a theoretical framework for improving the accuracy of teachers' judgments.

Finally, we will describe a professional development program designed to improve teachers' prediction accuracy.

1.1 Measuring the Accuracy of Teacher Judgments

Südkamp, Kaiser, and Möller (2012) recently conducted a meta-analysis of research on the accuracy of teachers' judgments of student academic achievement (see also (Hoge & Coladarci, 1989), and made a strong case for the importance of investigating teachers' judgment accuracy. They described multiple ways these judgments may affect student achievement. In particular, these judgments guide instruction (e.g., Alvidrez & Weinstein, 1999), identify struggling students (e.g., Bailey & Drummond, 2006), influence teachers' expectations about students' ability (e.g., Brophy & Good, 1986), shape feedback to students and parents (e.g., Hoge & Coladarci, 1989), and influence students' academic self-concept (e.g., Möller, Pohlmann, Köller, & Marsh, 2009).

In the literature, accuracy has been operationalized as the correlation between teachers' predictions of students' performance and students' actual performance. The literature suggests teachers vary widely in judgment accuracy (for reviews see Hoge & Coladarci, 1989; Südkamp et al., 2012). For instance, in their meta-analysis of 75 articles on accuracy of teacher judgments, Südkamp et. al. (2012) reported accuracy as high as .80 (Methe, Hintze, & Floyd, 2008) and as low as -.03 (Graney, 2008). Understanding the different levels of accuracy is complicated by the fact that researchers have used a variety of approaches to compute the correlation between predicted and actual performance.

The most common approach used to compute the correlation is to have teachers predict their students' achievement on a test, administer the test to students, and then simply stack the

data from all the teachers and compute the correlation between predicted and actual student performance (e.g., Begeny, Eckert, Montarello, & Storie, 2008; Freeman, 1993). This method computes the correlation across all the students disregarding the fact that students were nested within different teachers' classrooms. Although this will certainly provide a measure of the relation between predicted and actual student achievement, the correlation may be difficult to interpret because it could be influenced by group differences in achievement. For instance, if one teacher works with 25 high achieving students and another teacher works with 25 low achieving students, and the teachers' predictions reflect the group differences in achievement (i.e., the predicted and actual performance for the high achieving students are higher than those for the low achieving students), the correlation computed across these 50 students could be quite high—indicating highly accurate teacher judgments. However, this correlation could be quite high (due to group differences in achievement) even if the teachers do not accurately differentiate levels of student achievement within their own classes. Put differently, group differences in achievement could mask the true relationship between predicted and actual performance—when the teachers' data are combined and analyzed as a whole, ignoring that data came from independent teachers. To be fair, group differences may be mitigated when data from several teachers are combined. Nonetheless, any group differences that exist could inflate the correlation.

Another approach used to examine the accuracy of teachers' judgments of student achievement is to select one or two students from a classroom and have teachers predict performance for only that small subset of students in their classes, and then stack the data for all the teachers and compute the correlation across all students (e.g., Feinberg & Shapiro, 2009). Although this correlation represents the relation between predicted and actual student

achievement across teachers, it does not represent an individual teachers' ability to differentiate student performance within a classroom because the teacher only predicted performance for one or two students from his or her class. Given that effective teaching is influenced by a teachers' ability to differentiate instruction across students in a class (e.g., Donovan et al., 1999), measures of judgment accuracy should assess a teachers' ability to differentiate students who understand the material from those who do not, which would guide instruction (e.g., Alvidrez & Weinstein, 1999). Computing the correlation across teachers provides no information about how accurately teachers can judge differences in student skill levels, which raises questions about the value of this measure of judgment accuracy.

Ready and Wright (2011) modeled the nested nature of the data by using HLM to examine the relation between predicted and actual student achievement. They found a significant relation between predicted and actual achievement ($r = .60$ in the fall, $r = .66$ in the spring) even after accounting for the interdependence of students nested within teachers. They also identified student variables that were related to teachers' predictions (e.g., gender, race, and social class); however, none of these were as strongly related to teachers' predictions as students' prior test performance.

Finally, Helmke and Schrader (1987) examined teachers' judgment accuracy by computing a correlation for each teacher individually. That is, for each teacher, they computed an intra-individual correlation between fifth grade students' predicted and actual performance on a test of mathematical skills. They reported the judgment accuracy for all 32 teachers in the study, with accuracy ranging from approximately .05 to .90.

There are several advantages to using this intra-individual correlation as a measure of judgment accuracy. At a theoretical level, judgment accuracy is important because judgments about student learning guide instruction. Thus, accuracy should measure an individual teacher's ability to differentiate levels of learning among the students in his or her classroom, which is precisely what the intra-individual correlation provides.

At a more practical level, an advantage to measuring judgment accuracy at the individual teacher level is that it is possible to identify other teacher-level variables that could predict judgment accuracy (e.g., years of teaching experience, content knowledge, teaching practices). It is also possible to examine the relation between judgment accuracy and student outcomes (e.g., achievement). Perhaps most important in the context of the present investigation, in which professional development was delivered to improve judgment accuracy, having accuracy at the individual teacher level, made it possible to provide valuable feedback to teachers regarding their judgment accuracy, and evaluate the effectiveness of different aspects of the professional development (Day, 1999; de Vries, Jansen, & van de Grift, 2013).

1.2 Improving Judgment Accuracy

Teachers' judgments of students' academic performance are hypothesized to be important because these judgments guide subsequent instruction, which in turn influences student achievement. Thus, it follows that improving the accuracy of judgments should lead to more effective teaching, which should improve student achievement. The question is, "How can judgment accuracy be improved?"

To answer this question, we must first understand how teachers judge student learning. Brunswik (1956) suggested such judgments are made based on inferences one draws from

available cues. For example, a teacher could judge a student has learned the material based on a student's correct response to a question posed to the class. The teacher could also judge that the material has been learned because a student did not look confused or ask questions during the lesson. According to Brunswik (1956), the accuracy of judgments is determined in part by the *diagnosticity* of the cues used to make the judgment. Judgment accuracy will improve when the cues used to make a judgment are more diagnostic of subsequent student performance (for a more detailed discussion of how cue used affects judgment accuracy see (Dunlosky & Thiede, 2013; Koriat, 1997). In the example above, judgment accuracy should be better when the judgment is based on a correct response to a question than when the judgment is based on silence because the former is more diagnostic of learning than the latter.

There are a variety of cues available to judge student learning. The diagnosticity of a particular cue can change from one situation to the next (Thiede, Griffin, Wiley, & Anderson, 2010). For instance, a formative assessment aligned to a clearly defined learning objective should provide cues that are diagnostic of performance on a summative assessment, provided it is aligned to the same learning objective. In contrast, if the formative assessment is not aligned to the learning objective, it will not provide diagnostic cues. To improve the accuracy of teachers' judgments of student learning, teachers must (a) structure instruction to provide cues that are more diagnostic of the kind of learning that will ultimately be assessed on the summative assessment of learning, and then (b) use these more diagnostic cues to judge student learning.

Effective instruction requires a clearly defined learning objective, assessment of student learning during instruction (i.e., formative assessment), and adjustment of instruction to respond to students' level of learning (e.g., Black & Wiliam, 1998; Stiggins & Chappuis, 2006). To

effectively adjust instruction to facilitate student learning, teachers must be able to accurately judge student learning; however, accurate judgments of student learning may not produce greater student learning because accurate judgments do not necessitate the teacher will adjust instruction to address individual student's needs. That said, as accurate judgments are necessary but not sufficient, working to improve teachers' judgment accuracy is an important first component to improving instruction.

Our professional development was designed to help teachers teach mathematics in a student-centered way and to incorporate formative assessments into instruction (William, 2011). Focusing on cues produced from the student-centered aspect of formative assessments, which should be predictive of subsequent tests over the same material, was hypothesized to improve teachers' judgment accuracy.

1.3 Developing Mathematical Thinking

Developing Mathematical Thinking (DMT) is an instructional model that hinges on five key dimensions of classroom practice: focusing on students' initial ideas through problem solving, encouraging multiple solution strategies and models, pressing students conceptually, maintaining a focus on the structure of the mathematics, and addressing misconceptions (Brendefur, 2008; Brendefur, Thiede, Strother, Bunning, & Peck, 2013). DMT was chosen for this study because this multidimensional framework for teaching encourages teachers to monitor their students' mathematical ideas and to then modify their instruction based on what they observed from several perspectives. Here, we briefly describe the relevant research that frames this instructional process.

DMT is an instructional model built on the Dutch notions of "progressive formalization" and "mathematizing" (Freudenthal, 1973, 1991; Treffers, 1987) and on a socio-cognitive framework similar to Carpenter and Lehrer's (1999) elements of effective instruction. As Gravemeijer and van Galen (2003) describe, progressive formalization is a process of first allowing students to utilize informal strategies for solving contextual problems and ways to model these approaches, and then, by critically examining these informal strategies and models (monitoring students' learning), teachers press students to develop more sophisticated, formal, and abstract strategies and mathematical models. By examining the relationship among enactive, iconic and symbolic models (Bruner, 1964), teachers' modify instruction for students to learn which manipulations make sense for a given context and how to develop more generalizable procedures.

The DMT process starts with carefully chosen tasks – typically contextualized (Brendefur, Carney, Hughes, & Strother, In Press; Doerr, 2006; Larsen & Bartlo, 2009; Rosales, Vicente, Chamoso, Muñez, & Orrantia, 2012; Simon & Tzur, 2004). To solve these problems, students must model the situation to some degree. Rather than beginning with the standard algorithms and attempts to concretize them, teaching begins with students' commonsense solutions to contextual problems that are real for them. By reflecting on the solution procedures they have used, students develop and are introduced to more sophisticated models and procedures that they can also use in other situations (Gravemeijer & van Galen, 2003, p. 114).

Teachers, then, press students to make connections between existing knowledge (informal ideas) and new knowledge (more formal mathematical ideas) required by Hiebert and Carpenter's (1992) concept of conceptual understanding through solving novel problems. By

critically examining their own and others' strategies and models, students build a functional understanding, which exemplifies the importance of social interactions in classrooms. "By thinking and talking about similarities and differences between arithmetic procedures, students can construct relationships between them. … the instructional goal is not necessarily to inform one procedure by the other, but rather, to help students build a coherent mental network in which all pieces are joined to others with multiple links" (Hiebert & Carpenter, 1992, p. 68). This is an important aspect in teaching. Typically, teachers and students see mathematics as a series of procedures and definitions that build in complexity throughout the K-12 curriculum. But certain fundamental ideas or "structural components" appear continually throughout mathematics, whether at the primary grade level or later in high school courses. When instruction does not focus on the structure of mathematics, students often rely on memorized tricks or formulas and have difficulty solving complex problems or applying mathematics to new situations. The DMT framework embeds the language of the structural components within instruction and feedback to students on how to articulate and critique their own and others' mathematical models. Through these interlinked processes of modeling situations mathematically and analyzing different methods, teachers must know how each student is thinking about a problem and what pedagogical step to take next in order to press students to progressively formalize their ideas. This process of asking students to connect methods and generalize their thinking with questions or problem variations is an important part of a teacher's monitoring or formative assessment.

1.4 DMT Professional Development

In general, the DMT professional development takes place over one to three years and includes explicit and embedded professional development. The explicit professional

development includes a 45 hour course focusing on building participants' knowledge of students' thinking around mathematics (Ball, Hill, & Bass, 2005; Depaepe, Verschaffel, & Kelchtermans, 2013; Prescott, Bausch, & Bruder, 2013; Shulman, 1986), which includes knowledge of the strategies and/or models that students typically generate for given topics and how to press them to use particular models based on their thinking, knowledge of the multiple strategies and representations (i.e., enactive, iconic, and symbolic) possible for a given topic and the numerous connections that exist among them, and knowledge of how to utilize and/or sequence models (and tasks) to progressively formalize participants' thinking.

In addition to explicit professional development, there is embedded professional development through 6-10 in school meetings conducted throughout the year. This focus is on building the above aspects of knowledge both prior to and following a particular unit of instruction. Finally, 4-8 in class visits are conducted for the purpose of modeling instruction, co-teaching, monitoring fidelity of implementation and/or observing elements of the DMT framework in teachers' instruction.

The DMT professional development paralleled what we wanted to observe occurring in the classroom: provide a task or situation, allow for students to solve the task using any strategy of their choice, focus on modeling the situation, and then provide feedback through structure and misconceptions on how to improve their thinking and modeling techniques. Although the professional development covers many mathematical topics, we provide a typical example.

First, we purposely select a task that has multiple entry points for the participants and can be solved using a range of informal and formal models. One such task was Race Day:

> *Let's say that a race is taking place and you have just run 2 ½ miles and you are 1/3 of the way. How much farther do you have to run? Be prepared to explain your approach/strategy, how and why you modeled it the way you did, and whether your model is generalizable to other tasks. Then, explain what the problem type is, other ways students might solve the problem, and next possible instructional moves.*

Professional development participants begin working individually, but often quickly begin discussing approaches to solving the task in their small groups. In the meantime the facilitator walks around the room examining representations, facilitating small group discussions, and assisting individuals in creating their representations as needed. In particular, the facilitator identifies a range of participant-generated models for an upcoming class discussion. For this particular problem, the instructor looks for (1) informal iconic drawings, (2) 'educated' guess-and-check strategies, (3) tables, and (4) informal equations. The facilitator looks for a model that highlights an important idea and asks the participant or group members to recreate the model on poster paper for the whole-class discussion. These models are then posted around the room. When the facilitator does not see a particular model while monitoring participants' work, the facilitator looks to find a participant whose work and thinking connects to the 'missing' model and then presses the participant via questioning to generate the model based on his or her existing work.

The facilitator then encourages the participants to formalize their thinking by using a bar model or number line and equations to model their approach. Participants then practice explaining their approach to partners, while a few are asked to present their approach and how

one model afforded an efficient way of solving the problem or a way of deeply understanding the situation. Next, participants are given new situations that progress in difficulty and are asked to solve these problems using the more formal iconic and symbolic models. Finally, participants are asked to explain when the model they used should be used and when they should not be used. This discussion then incorporates pedagogical decisions regarding what they would do next in their classroom.

During this process the facilitator identifies a range of models across the enactive - iconic – symbolic trajectory (Bruner, 1964) in order to highlight the use of this framework as an instructional tool during the class discussion. The discussion typically begins with a focus on either an enactive or an informal iconic model of the task. Participants are asked to articulate their understanding of each model and are pressed to make conceptual connections that can be established between the models, with the goal of demonstrating how to use multiple participant-generated models to progressively formalize understanding.

By building these different aspects of teachers' knowledge we create shifts in their knowledge and practice. For example, at the beginning of the professional development, many participants while examining student thinking often focus on the correctness of a students' answer: "I'm surprised this student missed this problem." Later in the year, during the embedded professional development, teachers begin to shift their focus regarding student thinking and begin to use students' ideas to modify instruction: "This student decomposed the number incorrectly. Should I have them model it on a number line?" By increasing teachers' pedagogical knowledge, we shift what teachers focus on while teaching, which eventually shifts their instructional practices.

In sum, the DMT professional development is designed to promote student-based instruction of mathematics, which includes using formative assessment to estimate each student's learning trajectory. We hypothesize that this professional development will increase teachers' knowledge of individual students' thinking, which are cues that should be predictive of students' subsequent test performance. Thus, we predict that judgment accuracy will be greater for teachers who have been through DMT professional development than for those who have not. Furthermore, as judgment accuracy is hypothesized to inform instruction, which influences student learning, we also predict that student achievement will be greater for teachers who have been through the DMT professional development (and implement it with high fidelity) than for those who have not.

1.5 Overview of Studies

This investigation consisted of two studies. We wanted to examine the effect of DMT professional development on teachers' judgment accuracy. As described above, we hypothesized that DMT would increase judgment accuracy; however, the degree to which DMT was implemented was expected to influence accuracy. In particular, we expected judgment accuracy to be greater for teachers who implemented DMT with high fidelity than for those who implemented DMT with low fidelity.

In Study 1, participants were teachers in Grade 3 – 5, who had participated in three years of DMT professional development. Based on classroom observations and interviews, teachers were classified as either high- or low-fidelity implementers (see below for more details). High-fidelity implementers were those who (a) pressed students to explore the benefits and limitations of strategies and models used to solve problems, (b) focused on developing students

understanding of various approaches to solving problems, (c) used mistakes and misconceptions as valuable tools to build mathematical understanding, (d) used tasks and activities that encouraged analysis and discussion of fundamental mathematical components related to the topic being studied, and (e) valued and built upon students' solutions to problems and provided opportunities for students to share their approach with others. Those who were low-fidelity implementers were those who (a) focused on the transmission of knowledge to students, often included rote repeated practice, rather than the construction of knowledge by students, (b) focused on providing one correct procedure or method for students to learn, (c) addressed mistakes and misconceptions by a thorough explanation of the correct process or procedure for solving a problem, (d) emphasized superficial elements of the mathematics, and (e) had students working individually, with the textbook or other resource materials directing the daily instructional topic and practice problems. As implementation of DMT provides teachers with more diagnostic cues for judging student learning, we hypothesized monitoring accuracy would be greater for the high-fidelity implementers than for the low-fidelity implementers. We sought to answer the question, is judgment accuracy affected by the fidelity of implementation of DMT?

In Study 1, all participating teachers had received DMT professional development; therefore, this study did not provide a comparison of teachers who received DMT professional development and those who did not. In Study 2, we sought to answer the question, is judgment accuracy affected by DMT professional development?

In Study 2, participants were teachers in Grades 2 – 5, from eight schools. Four schools were randomly assigned to participate in the DMT professional development, and the other four were a control group. Again, as implementation of DMT provides teachers with more diagnostic

cues for judging student learning, we hypothesized monitoring accuracy would be greater for the DMT teachers than for the non-DMT teachers.

2. Study 1

2.1 Method

2.1.1 Participants and design

In this study, participants were 36 Grade 3 – 5 teachers who had been in the DMT professional development for three years. As part of the professional development, teachers' mathematics instruction was observed between 3 and 5 times each year. Four researchers (two which are authors) used the DMT observation instrument to rate teachers on the five dimensions of effective teaching (Brendefur, 2008; Carney, Brendefur, Thiede, & Hughes, 2014; Hiebert et al., 1997): (a) taking students' ideas seriously, (b) pressing students conceptually, (c) encouraging multiple representations/models, (d) addressing misconceptions, and (e) focusing on the structure of mathematics. The five dimensions have been shown to be reliable, with Cronbach's alphas between .83 and .94. Observation scores on the five dimensions and interviews with teachers about mathematics instruction and student learning were used to classify teachers as high-fidelity implementers of DMT or low-fidelity implementers of DMT. Teachers who were consistently high on all five dimensions were classified as high-fidelity implementers; whereas, teachers who were consistently low on all five dimensions were classified as low-fidelity implementers. The inter-rater reliability of the classifications was quite high (Kappa = .94). Fifteen teachers (evenly distributed across grades) were classified as high-fidelity implementers and 21 (evenly distributed across grades) were classified as low-fidelity

implementers. Thus, group (high-fidelity implementers versus low-fidelity implementers) was a between-subjects variable.

2.1.2 Measures

The key outcome variable for this investigation was accuracy of teachers' prediction of student learning, which we measured as in Helmke and Schrader (1987). That is, prediction accuracy was operationalized as the intra-individual correlation between a teacher's predictions of student performance and the students' actual performance—computed across the students within a class. We used a non-parametric, Goodman-Kruskal gamma correlation[i] because Nelson (1984) recommended using this measure of association for these kinds of data.

As noted in the introduction, prediction accuracy is hypothesized to be important because it informs decisions related to instruction, which in turn influences student achievement. Thus, more accurate monitoring should lead to better differentiation of instruction, which should produce greater gains in student learning.

2.1.3 Procedure

In the spring, teachers previewed a ten-item mathematics test designed to assess students' knowledge of number sense (e.g., operations, patterning and sequencing, reasoning about quantity). After previewing the test, a teacher predicted the performance of each student in his or her class on the test. The teacher then administered the test, which was later scored by the researchers.

2.2 Results

Predicted and Actual Test Performance. Prediction accuracy was operationalized as the intra-individual correlation between predicted and actual performance. To be complete, we first

report descriptive statistics on predicted and actual performance. For each teacher, we computed the mean predicted and actual performance across the students. We then computed the mean predicted and actual performance across the teachers in each group. As seen in Table 1, predicted performance was greater for high-fidelity implementers than for low-fidelity implementers, $t(34) = 2.38$, $p = .02$. Actual performance was also greater for high-fidelity implementers than for low-fidelity implementers, $t(34) = 5.05$, $p < .01$, *Cohen's d* = 1.73.

Insert Table 1 here

Prediction Accuracy. For each participating teacher, we computed a gamma correlation between a teacher's predictions of student performance and the student's actual performance. The mean of the intra-individual correlations was then computed across teachers within each group. Mean prediction accuracy was significantly greater for high-fidelity implementers than for low-fidelity implementers, $t(34) = 4.35$, $p < .001$, see Figure 1.

Insert Figure 1 here

Prediction Accuracy and Student Achievement. Given that prediction accuracy should inform instruction, and in turn student learning, we expected to observe a positive correlation between prediction accuracy and student achievement. Indeed, we found a moderate positive correlation between these variables ($r(36) = .56$, $p < .05$). Although the significant correlation does not imply causation, it is consistent with the notion that accurate teacher judgments can

provide better information for differentiating instruction, which then produces better student achievement.

2.3 Discussion

Judgment accuracy is influenced by the cues used to make judgments (Brunswik, 1956; Koriat, 1997). Accuracy will improve when cues used to make judgments are more diagnostic of subsequent performance. Based on classroom observations, teachers who implement DMT with high fidelity use more diagnostic cues focused on students' thinking and mathematical understanding than teachers who do not; thus, we hypothesized that accuracy would be greater for high-fidelity implementers than for low-fidelity implementers. Findings from Study 1 are consistent with our hypothesis.

One could make a case that this study simply showed that effective teachers of mathematics more accurately monitor their students' learning—because teachers were in groups based on observations of their instructional practice. To speak to the causal role judgment accuracy plays in student learning, it would have been more compelling to have selected teachers based on judgment accuracy and looked for difference in instructional practice and student learning. [Thiede (1999), showed the importance of attending to the causal relation in drawing conclusions about the importance of monitoring accuracy in learning]. However, the results of this study suggest judgment accuracy could be an important variable to investigate further.

In Study 2, we compared monitoring accuracy of teachers who participated in the DMT professional development to those who did not. Moreover, schools were matched on previous mathematics performance, and teachers were randomly selected to participate in the professional development; therefore, teacher effectiveness should not have differed dramatically across the

DMT group and the control group. The participating district for this study (which was different than the district in Study 1) was actively preparing for the transition to assessments, which are more aligned to international tests like TIMSS and PISA; therefore, we examined the accuracy of teachers' predictions of student performance on a computational test and a conceptual test (as had been conducted in Helmke & Schrader, 1987). As DMT professional development was thought to focus teachers on more diagnostic cues related to student learning of mathematics, we hypothesized that prediction accuracy would be greater for the DMT group than the control group.

3. Study 2

Study 2 was conducted as part of large federally funded project to examine a professional development program designed to improve the accuracy of teachers' judgments of student achievement by improving their use of formative assessments as a means to monitory learning.

3.1 Method

3.1.1 Setting, Design, and Participants

This study was set in large school district in Southeastern Idaho. Prior to recruiting schools to participate in the study, the study was described to all the elementary school principals in the district, and interested principals (N = 24) volunteered to participate in the study. The district selected eight schools to participate in the study and created two matched sets of schools based on prior student achievement, SES, and racial composition. Prior to assigning the schools to a condition, the principals were asked to select two teachers from each grade to participate in the DMT professional development. We randomly selected schools to receive the professional

development or serve as part of a control group. Control schools were assigned to the professional development group the following year.

As in Study 1, the key dependent variable was the accuracy of teacher predictions (the intra-individual correlation between teachers' predictions of student mathematics performance and actual student performance). We computed two intra-individual correlations for each teacher: one for a test of computational skill and another for a test of conceptual knowledge. Thus, Test (Computational versus Conceptual) was an independent variable in the study. Group (DMT versus No-DMT) was the other independent variable in this study.

We again conducted an exploratory analysis to evaluate the hypothesized relation between prediction accuracy and student achievement. In this study, student achievement was assessed with the mathematics section of the Measures of Academic Progress (MAP). The MAP test was administered to students starting at Grade 1; thus, we had change scores available for students in grades 2 through 5.

Participants were 64 Grade 2 through 5 teachers: 32 in the DMT group and 32 in the non-DMT group. As seen in Table 2, the groups did not differ in years of teaching experience, pretest knowledge of mathematics pedagogy and content (measured with the Elementary School Number Concepts and Operations 2001, (Learning Mathematics For Teaching, 2005), or pretest knowledge of classroom assessment (measured with the Classroom Assessment Literacy Inventory, (Impara, 1993), all *t*s < 1. The groups were also equal in education level—both had 11 teachers with a master's degree and 21 with a bachelor's degree, $\chi^2(1) = 0$. Thus, the groups were equivalent on key variables prior to the professional development.

Insert Table 2 here

3.1.2 Procedure

Early in the school year, teachers completed all pretest measures, which included a survey of demographics, the measure of knowledge of mathematics pedagogy and content, and the measure of assessment knowledge. Teachers in both groups received 15 hours of professional development focused on student-centered instruction, knowledge for teaching, assessing knowledge consistent with levels of reasoning, and using formative assessments to track students' learning. The DMT professional development included regular visits to teachers' classrooms for observation, coaching, and co-teaching of lessons and included 6-8 planning sessions prior to and following a mathematics unit. In the pre unit-planning sessions, grade-level teachers met to analyze summative assessments for the unit, select or develop formative assessment items to assess progress in the unit, and specific strategies for providing student-centered instruction of the mathematical concepts (for more details, see Brendefur et al., 2015). The post unit-planning sessions involved examining the results of the formative and summative assessments. Here, teachers examined their student thinking and created next steps for instruction.

In April, teachers completed the prediction task. They previewed the tests of computational skill and conceptual knowledge of number sense (e.g., operations, patterning and sequencing, reasoning about quantity)—the tests had good internal consistency (Cronbach's alpha = .88 and .84, respectively). They then predicted how many items out of five each student

would correctly answer on each test. The teachers then administered the tests and returned them to the researchers for scoring.

3.2 Results

Predicted and Actual Test Performance. As prediction accuracy is the relation between teachers' predictions of student performance and students' actual performance on the mathematics tests, we report these variables first. We conducted a 2 (Group: DMT versus No-DMT) x 2 (Test: Computational versus Conceptual) mixed effects analysis of variance (ANOVA) to compare both predicted and actual performance. As seen in Table 3, predicted performance did not differ across groups, $F(1, 62) = .03$, $MSe = .69$, $p = .87$. Predicted performance was greater for the test of computational skill than for the test of conceptual knowledge, $F(1, 62) = 100.97$, $MSe = .37$, $p < .001$, *partial eta squared* = .62. The interaction was not significant, $F(1, 62) = .42$, $MSe = .37$, $p = .52$. The pattern of results was the same for students' actual performance. We had missing data for three teachers (two in the DMT group and one in the No-DMT group). As seen in Table 3, actual performance did not differ across groups, $F(1, 59) = 1.88$, $MSe = .61$, $p = .18$. Actual performance was greater for the test of computational skill than for the test of conceptual knowledge, $F(1, 59) = 135.83$, $MSe = .41$, $p < .001$, *partial eta squared* = .70. The interaction was not significant, $F(1, 59) = .13$, $MSe = .41$, $p = .72$.

Insert Table 3 here

Prediction Accuracy. As in Study 1, prediction accuracy was operationalized as the intra-individual correlation between predicted and actual performance. Each teacher had two measures

of prediction accuracy: one for the test of computational skill and one for the test of conceptual knowledge. The three teachers with missing student performance had indeterminate correlations. As seen in Table 4, prediction accuracy was greater for the DMT group than for the No-DMT group, $F(1, 59) = 7.58$, $MSe = .10$, $p = .008$, *partial eta squared* = .11. Prediction accuracy was greater for the computational test than for the conceptual test, $F(1, 59) = 73.01$, $MSe = .11$, $p < .001$, *partial eta squared* = .55. The interaction was not significant, $F(1, 59) = 1.17$, $MSe = .11$, $p = .28$. These findings are consistent with our prediction that judgment accuracy would be greater for teachers who participate in the DMT professional development than for those who do not.

Prediction Accuracy and Student Achievement. As in Study 1, we conducted an exploratory analysis to examine the relation between prediction accuracy and change in student achievement (as measured by the mathematics section of the MAP). We found accuracy for predicting computational performance was significantly correlated with change in achievement ($r(64) = .32$, $p < .05$); whereas, accuracy for predicting conceptual performance was not correlated with change ($r(64) = .16$, $p > .05$). Thus, accuracy of predictions of computational performance is significantly correlated to student performance on a standardized test that is largely computational.

3.3 Discussion

A key finding from this study was that prediction accuracy was greater for teachers who participated in the DMT professional development than for teachers who did not. The superior accuracy held across tests of computational skill and conceptual knowledge. Helmke and Schrader (1987) also attempted to assess the accuracy of teachers' predictions across computational and conceptual tests; however, their test of conceptual knowledge was not

reliable; therefore, they reported only accuracy of predictions for a computational test. In this study, we had a reliable test of conceptual knowledge, and although accuracy was greater for the DMT teachers than for the no-DMT teachers, both groups struggled to accurately predict performance on a test of conceptual knowledge. In light of the increased focus on conceptual understanding and application of mathematics internationally, this may be problematic. The low levels of prediction accuracy suggest that teachers may not be able to accurately identify students in need of additional instruction, which could have a detrimental effect on student learning.

4. General Discussion

Effective instruction requires accurate monitoring of student learning. This information is crucial to making informed decisions about whether educational objectives have been reached by all students, and if not, which students need additional help to reach the objectives (Donovan et al., 1999; Glaser, Chudowsky, & Pellegrino, 2001). Thus, it is important to understand ways to assess the accuracy of teachers' monitoring of student learning, and to improve the accuracy of teachers' monitoring.

In the present research, we operationalized judgment accuracy as an intra-individual correlation between predicted and actual performance (an approach widely used in metacognitive research, Dunlosky & Thiede, 2013). An advantage to this approach is it provides a measure of accuracy for each teacher, which makes it possible to compare accuracy across individuals or groups of teachers. Thus, researchers can gain insights into what factors might affect prediction accuracy by simply examining the relation between variables and prediction accuracy. For instance, we can examine what instructional practices lead to better prediction accuracy (as in

Study 1), which can help us develop new testable hypotheses about what factors might influence accuracy, or help us develop new interventions to improve accuracy (as in Study 2).

Brunswik (1956) developed a promising theoretical framework to help understand factors that affect monitoring. This theory states that monitoring involves making inferences about reality and accuracy is driven be the cues used to make a judgment. This theory suggests accuracy will improve as cues used to make judgments become more diagnostic of the reality or, in our case, more focused on students' actual understanding than on the content taught. Koriat (1997) developed the cue-utilization framework of monitoring, which also suggests monitoring accuracy is influenced by the cues used to make judgments. The cue-utilization framework, which has been cited more than 300 times, revolutionized the way researchers have approached metacognitive monitoring. Moreover, focusing on the cues used to make judgments has led to help identifying more diagnostic cues for making metacognitive judgments (e.g., Thiede et al., 2010) and the development of new interventions to improve monitoring accuracy (e.g., Redford, Thiede, Wiley, & Griffin, 2012).

The cue-utilization framework guided the present research. We hypothesized that formative assessment practices, such as analysis of student thinking through the use of exit tickets, provide teachers with diagnostic cues for predicting students' performance on an upcoming test; therefore, we expected prediction accuracy would be great for teachers who use more formative assessment practices than for teachers who use less. The results from Studies 1 and 2 were consistent with this hypothesis. In Study 2, judgment accuracy was greater for teachers who participated in professional development that promoted use of formative assessments in mathematics instruction than for teachers who did not. In Study 1, judgment

accuracy was greater for teachers who implemented the professional development with high fidelity than for teachers who implemented it with low fidelity.

With the increased emphasis on accountability in various countries, teachers are expected to accurately evaluate their students' understanding. Many schools, including those where this research was conducted, have participated in professional development to improve instructional practice around use of formative assessments (e.g., Chappuis, Chappuis, & Stiggins, 2009). This professional development presumably improves instructional practice; however, it is unclear whether it improves judgment accuracy—especially within a particular domain. In contrast to this professional development that targets use of formative assessment more generally, DMT professional development is focused on improving mathematics instruction and monitoring to students' learning of mathematics. Thus, DMT focuses teachers on cues specific to monitoring students' learning of mathematics.

This is the first study to compare the accuracy of teachers' predictions of student performance of computational skill and conceptual knowledge. Results from Study 2 showed teachers (in both the DMT and no-DMT group) more accurately predicted computational skill than conceptual knowledge. Moreover, accuracy levels were quite low for predicting conceptual knowledge. Poor monitoring accuracy occurs because people use cues that are not diagnostic of subsequent test performance (Benjamin, Bjork, & Schwartz, 1998; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Thiede et al., 2010) or they ignore cues that are diagnostic of test performance (Koriat & Bjork, 2005). In the context of the present investigation, teachers either did not have access to diagnostics cues while predicting conceptual knowledge or if they had access, they did not utilize those cues. With the increased focus internationally on conceptual

knowledge imbedded within learning standards, it will be important to conduct additional research to better understand why teachers may struggle to predict conceptual knowledge.

The present research suggests that DMT professional development improves judgment accuracy by focusing teachers on appropriate cues for judging student learning. More work is needed to identify cues that are predictive of conceptual understanding. Once these cues are identified, we can work to help teacher use these cues for judging their students' conceptual understanding—and adjusting their instruction to improve student learning.

[i] Nelson (1984) recommended using a Goodman-Kruskal gamma correlation (Goodman & Kruskal, 1954) for these kinds of data. Gamma is computed by examining the direction of one variable relative to another. If one variable (e.g., metacomprehension judgment) is increasing from one text to another and the other variable (e.g., test performance) is also increasing across this same pair of texts, this is considered a concordance (C). By contrast, if one variable is increasing from one text to another and the other variable is decreasing across this same pair of texts, this is considered a discordance (D). Concordance and discordance is computed across all pairs of items. The total number of each is used to compute the correlation coefficient, Gamma = (C – D)/(C + D). Once computed, the Gamma coefficients then represent a continuous and normally distributed measure of judgment-performance correspondence suitable for analysis with most GLM approaches.

References

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731.

Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment, 11*(3-4), 149-178.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade and how can we decide? *American Educator*, 14-46.

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education, 5*(1), 7-74.

Brendefur. (2008). Connecting elementary teachers' mathematical knowledge to their instructional practices. *The Researcher, 21*(2), 1-18.

Brendefur, J. L., Carney, M. B., Hughes, G., & Strother, S. (In Press). Framing professional development that promotes mathematical thinking. In E. Ostler (Ed.), *Emerging trends and perspectives in STEM learning.*

Brendefur, J. L., Thiede, K., Strother, S., Bunning, K., & Peck, D. (2013). Developing mathematical thinking: Changing teachers' knowledge and instruction. *Journal of Curriculum and Teaching, 2*(2).

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 328-375). New York: Macmillan.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19*(1), 1-15.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*: Univ of California Press.

Burkhardt, H., Schoenfeld, A., Abedi, J., Hess, K., & Thurlow, M. (2012). Content Specifications for the Summative assessment of the Common Core State Standards for Mathematics (pp. 1-145): Smarter Balanced Assessment Consortium.

Carney, M. B., Brendefur, J. L., Thiede, K., & Hughes, G. (2014). *DMT teacher observation instrument*. Paper presented at the Annual meeting of the National Council of Teachers of Mathematics, New Orleans, LA.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. Romberg (Eds.), *Mathematics Classrooms that Promote Teaching for Understanding* (pp. 19 - 32). Mahwah, NJ: Lawerance Erlbaum Associates.

Chappuis, S., Chappuis, J., & Stiggins, R. (2009). The quest for quality. *Educational Leadership, 67*, 14-19.

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*: Psychology Press.

de Vries, S., Jansen, E. P., & van de Grift, W. J. (2013). Profiling teachers' continuing professional development and the relation with their beliefs about learning and teaching. *Teaching and Teacher Education, 33*, 78-89.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12-25.

Doerr, H. M. (2006). Examining the tasks of teaching when using students' mathematical thinking. *Educational Studies in Mathematics, 62*(1), 3-24.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*: National Academies Press.

Doyle, W. (1977). Learning the classroom environment: an ecological analysis. *Journal of Teacher Education, 28*(6), 51-55.

Duffy, G. G., Miller, S., Parsons, S., & Meloth, M. (2009). Teachers as metacognitive professionals. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 240-256). New York, NY: Routledge.

Dunlosky, J., & Thiede, K. W. (2013). Metamemory. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 283-298). Oxford, England: Oxford University Press.

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*(6), 453-462.

Freeman, J. G. (1993). Two factors contributing to elementary school teachers' predictinos of students' scores on the Gates-MacGinitie reading test, Level D. *Perceptual and motor skills, 76*(2), 536-538.

Freudenthal, H. (1973). *Mathematics as an educational task*: Springer.

Freudenthal, H. (1991). Revisiting Mathematics Education: China Lectures.

Glaser, R., Chudowsky, N., & Pellegrino, J. W. (2001). *Knowing What Students Know:: The Science and Design of Educational Assessment*: National Academies Press.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications*. *Journal of the American Statistical Association, 49*(268), 732-764.

Graney, S. B. (2008). General education teacher judgments of their low-performing students' short-term reading progress. *Psychology in the Schools, 45*(6), 537-549.

Gravemeijer, K., & van Galen, F. (2003). Facts and algorithms as products of students' own mathematical activity. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 114-122). Reston, VA: NCTM.

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education, 3*(2), 91-98.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 22.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-97). New York: Macmillan.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Wearne, D., & Murray, H. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313.

Impara, J. C. (1993). Educational Administrators' and Teachers' Knowledge of Classroom Assessment. *Journal of School Leadership, 3*(5), 510-521.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187.

Larsen, S., & Bartlo, J. (2009). The role of tasks in promoting discourse supporting mathematical learning. In L. Knott (Ed.), *The role of mathematics discourse in producing leaders of discourse* (pp. 77-98). Charlotte, NC: Information Age Publishing.

Learning Mathematics For Teaching. (2005). Mathematical Knowledge for Teaching Measures: Number and Operations Knowledge of Students and Content. Ann Arbor, MI.

Methe, S. A., Hintze, J. M., & Floyd, R. G. (2008). Validation and decidion accuracy of early numeracy skill indicatiors. *School Psychology Review, 37*(3).

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129-1167.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological bulletin, 95*(1), 109.

NGA, & CCSSO. (2011). Common Core State Standards for Mathematics. 2010, from http://www.corestandards.org.

Prescott, A., Bausch, I., & Bruder, R. (2013). TELPS: A method for analysing mathematics pre-service teachers' Pedagogical Content Knowledge. *Teaching and Teacher Education, 35*, 43-50.

Ready, D. D., & Wright, D. L. (2011). Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities The Role of Child Background and Classroom Context. *American Educational Research Journal, 48*(2), 335-360.

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction, 22*(4), 262-270.

Rosales, J., Vicente, S., Chamoso, J. M., Muñez, D., & Orrantia, J. (2012). Teacher–student interaction in joint word problem solving. The role of situational and mathematical knowledge in mainstream classrooms. *Teaching and Teacher Education, 28*(8), 1185-1195.

Shavelson, R. J. (1978). Teachers' Estimates of Student'States of Mind'and Behavior. *Journal of Teacher Education, 29*(5), 37-40.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Simon, M., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning, 6*(2), 91-104.

Stiggins, R. J., & Chappuis, J. (2006). What a difference a word makes: Assessment "for" learning rather than assessment "of" learning helps students succeed. *Journal of Staff Development, 27*, 10-14

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6*(4), 662-667.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*(4), 331-362.

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction – The Wiskobas Project*. Reidel: Dordrecht, The Netherlands.

William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree.

Author Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through R305A120265 to Keith Thiede, Jonathan Brendefur, Richard Osguthorpe, Michele Carney, and Jennifer Snow. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Figure Captions

Figure 1. Mean prediction accuracy for Study 1. Error bars are the standard error of the mean.

Table 1. Predicted and Actual Test Performance by Group

Group	Predicted Performance	Actual Performance
High-fidelity implementers (N = 15)	5.68 (.35)	4.92 (.39)
Low-fidelity implementers (N = 21)	4.20 (.51)	2.53 (.29)

Note. The numbers in parentheses are the standard error of the mean.

Table 2. Mean Years of Experience and Pretest Knowledge Scores

Group	Years of Experience	Math Pedagogical Knowledge	Assessment Knowledge
DMT (N = 32)	14.52 (1.55)	2.81 (.14)	.34 (.14)
No-DMT (N = 32)	13.14 (1.40)	3.01 (.08)	.43 (.15)

Note. The numbers in parentheses are the standard error of the mean.

Table 3. Predicted and Actual Test Performance by Group

Group	Computational Skill	Conceptual Knowledge
Predicted Performance		
DMT	3.46 (.14)	2.31 (.16)
No-DMT	3.36 (.10)	2.35 (.12)
Actual Performance		
DMT	3.13 (.10)	1.74 (.17)
No-DMT	2.89 (.11)	1.59 (.13)

Note. Numbers in parentheses are the standard error of the mean.

Table 4. Prediction Accuracy by Group

Group	Computational Skill	Conceptual Knowledge
DMT	.66 (.02)	.20 (.07)
No-DMT	.56 (.04)	-.02 (.08)

Note. Numbers in parentheses are the standard error of the mean.

Figure 1.