

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337874869>

# A Comparison of Perceived and Observed Learning from Professional Development: Relationships Among Self-Reports, Direct Assessments, and Teacher Characteristics

Article in *Journal of Teacher Education* · December 2019

CITATIONS

0

2 authors:



**Yasemin Copur-Gencturk**  
University of Southern California

12 PUBLICATIONS 108 CITATIONS

[SEE PROFILE](#)



**Ian Thacker**  
University of Southern California

7 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



CAREER: Development of Pedagogical Content Knowledge in Mathematics Among Beginning Teachers [View project](#)



Advancing Middle School Teachers' Understanding of Proportional Reasoning for Teaching [View project](#)

# **A Comparison of Perceived and Observed Learning from Professional Development: Relationships Among Self-Reports, Direct Assessments, and Teacher Characteristics**

## **Abstract**

The success of professional development programs has typically been determined based on their impact on teacher learning, without much attention being given to the data sources used. Large-scale studies have generally relied on teachers' self-reports, whereas small-scale studies have included more direct assessments and observations of teacher learning. The purpose of this study was to compare teachers' self-reported gains in mathematical knowledge for teaching with those measured by direct assessments. Quantitative analyses of the data collected from 545 teachers who participated in content-focused professional development programs indicated a lack of correlation between teachers' self-reports and direct assessments of their knowledge gains. Furthermore, different teacher-related factors were associated with the learning reported by these two measures. These findings speak to the need to pay careful attention to the outcome measures used to evaluate teachers' learning.

## Introduction

Professional development is a means of equipping in-service teachers with the necessary knowledge and skills to provide quality instruction and enhance their students' learning. Although a majority of teachers have reported participating in professional development activities (Banilower et al., 2013), empirical evidence is mixed regarding what kinds of learning opportunities are most effective in enhancing teachers' knowledge or improving their instructional practices, and in turn their students' learning (cf. Blank, las Alas, & Smith, 2008; Garet et al., 2011, 2016; Santagata, Kersting, Givven, & Stigler, 2011; Yoon, Duncan, Lee, & Scarloss, 2007). We argue that the paucity of conclusive evidence on professional development is partly due to the potential lack of alignment in what is captured by the different methods used to determine the success of a program. Specifically, the large-scale studies that played vital roles in identifying key features of effective professional development were based on teachers' self-reports, whereas current research on the effectiveness of professional development usually utilizes direct measures of teachers' learning. It is possible that teachers' perceptions of their learning might not capture the same construct as their learning measured by direct assessments, which, in turn, hinders our efforts to understand which conditions and according to which sources professional development programs seem to be effective.

Although some research has been done to provide evidence on the validity of self-reports, teachers' self-reports are used differently in the evaluation of many professional development programs. More explicitly, studies have examined the alignment of teachers' self-reports with direct assessments of the *current state* of a phenomenon (e.g., Kaufman, Stein, & Junker, 2016), rather than the alignment of teachers' self-assessments of their *improvement*, with the improvement detected by direct assessments. The former requires teachers to focus only on their

current knowledge or skills, whereas the latter requires teachers to compare their levels of knowledge and skill before and after participating in professional development. In many professional development studies utilizing self-reported data, teachers are typically asked to evaluate the change in their knowledge or practices (Garet, Porter, Desimone, Birman, & Yoon, 2001; Heck, Banilower, Weiss, & Rosenberg, 2008; Ingvarson, Meiers, & Beavis, 2005; Zwart, Wubbels, Bergen, & Bolhuis, 2009). Therefore, research is needed that explores the relationship between teachers' learning captured by self-reports versus direct assessments. The availability of valid teacher knowledge assessments now makes it possible to explore this phenomenon at a large scale. The purpose of this study was twofold: one was to compare teachers' self-reported knowledge improvements, among 545 teachers in 24 different content-focused, yearlong professional development programs, with those detected by a direct instrument designed to capture the same knowledge gains and the second was to investigate what teacher characteristics and practices were associated with teachers' learning according to which measure was used.

In the following section, we discuss how we conceptualized teachers' learning, namely, as the change in their mathematical knowledge for teaching (MKT). We then elaborate on why we expect discrepancies in perceived versus assessed learning.

### **Content Knowledge Needed in Teaching**

In this study, we focused on teachers' MKT because, both theoretically and empirically,<sup>1</sup> it is an important construct for teaching and student learning. As emphasized by Shulman (1986)

---

<sup>1</sup> Although recent studies on the role of teachers' MKT in students' mathematics achievement have yielded mixed evidence (cf. Hill, Rowan, & Ball, 2005; Kersting et al., 2012; Ottmar, Rimm-Kaufman, Larsen, & Berry, 2015; Rockoff, Jacob, Kane, & Staiger, 2011), the role of such knowledge in the quality of mathematics instruction is more salient (Author, 2015; Hill et al., 2008), suggesting that teachers with strong MKT are able to make the concepts students are expected to learn explicit and make fewer errors in their teaching.

three decades ago, teachers not only need to know the concepts they are expected to teach; they also need to know discipline-specific pedagogy and to be able to understand how their students learn concepts. Ball, Hill, and colleagues (e.g., Ball, Thames, & Phelps, 2008; Hill, Schilling, & Ball, 2004) elaborated further on what it means for teachers to know mathematics for teaching. To do so, they analyzed the work of teaching mathematics and created assessments to capture some aspects of this knowledge. The studies conducted by the instrument developers to determine the construct validity of the developed items indicated that teachers drew on their knowledge of mathematics and their knowledge of students' understanding of mathematics when they answered these questions (e.g., Hill, Dean, & Goffney, 2007).

In this study we aimed to understand the relationship between teachers' perceived knowledge improvement and the improvement captured by the direct assessments, and thus we focused on the kinds of knowledge teachers seem to draw on while answering the items in the MKT assessments. As such, we asked teachers to report the change in their knowledge of mathematics and knowledge of students' understanding of mathematics so that both data sources (self-reports and direct assessments) would focus on the same aspects of content knowledge for teaching.

### **Self-Reports on Learning**

Self-reports are widely used in behavioral, psychological, and medical research and have been shown to accurately reflect individuals' demographic information, emotions, self-efficacy, and interest (for reviews, see Chan, 2009; Stone, Bachrach, Jobe, Kurtzman, & Cain, 2000). Likewise, in educational research, teachers' reports on the frequency of their instructional practices appear to align with observations and interviews (Kaufman et al., 2016; Mayer, 1999; Ross, McDougall, Hogaboam-Gray, & LeSage, 2003), especially if teachers are asked to report

on their practices for a single class or for limited time frames (Newfield, 1980; Porter, Kirst, Osthoff, Smithson, & Schneider, 1993). However, studies have found discrepancies between self-reports and direct observations regarding the quality of instructional practices (Kaufman et al., 2016; Mayer, 1999). This is consistent with research in health, public policy, and other domains which find that individuals tend to be more accurate when reporting discrete events that are framed in terms of recent and distinct timelines and are less accurate when reporting on attitudes, attributes, or behaviors that are socially valued or disapproved (for reviews, see e.g., Bradburn, 2000; Tourangeau, Rips, & Rasinski, 2000).

Additionally, as mentioned, in many professional development studies, teachers are asked to report improvements in their knowledge or practices (e.g., see Jayanthi, Gersten, Taylor, Smolkowski, & Dimino, 2017; Sitzmann, Ely, Brown, & Bauer, 2010), which is different from asking teachers to report their practices for a limited time frame. When learners are asked to report increases in their knowledge, they must compare their current knowledge with a model of their previous understanding through self-reference, leaving many opportunities for inaccuracy. A body of research in experimental psychology suggests that individuals generally tend to make inaccurate judgments of their own learning (Dunlosky & Nelson, 1992; Koriat, 1997) and performance (i.e., people are not well calibrated; Glenberg & Epstein, 1987; Pieschl, 2009; Schraw, 2009; Schraw, Kuch, & Gutierrez, 2013; Stone, 2000), particularly when lacking basic competencies (Dunning, Heath, & Suls, 2004; Dunning, 2011; Kruger & Dunning, 1999). Although few studies, if any, have been conducted to compare teachers' self-reported learning during professional development with that captured by direct assessments, in a meta-analysis of 137 adult education and workplace-training studies that used self-reported knowledge as a major outcome, Sitzmann and colleagues (2010) found no correlation, on average, between self-

reported *increases* in knowledge and actual “cognitive” knowledge gains (weighted mean  $r = .00$  across  $k = 25$  studies, 95% CI =  $-.12$ – $.12$ ).

Generally, these studies suggest that children and adults are overconfident in their comprehension of newly learned skills, potentially because rapid training appears to promote skill acquisition and self-confidence but not necessarily the retention of skills (e.g., see Dunning et al., 2004). However, much of this evidence is based on students’ and adults’ evaluations of their own learning in controlled laboratory environments, which may be quite different from teacher learning in dynamic professional development environments. Notably, given that teachers’ classroom experience focuses on monitoring students and their knowledge acquisition, it may be the case that teachers are more accurate in their evaluations of their own learning when compared with people in other professions.

### **Potential Teacher-Related Factors Associated with Teacher Learning**

Regardless of the perceived learning captured by self-reports or the observed learning assessed by direct observations, more research is needed to understand the extent to which teachers’ characteristics and practices are associated with their learning during professional development. Research has identified several factors that predict peoples’ learning, including disciplinary expertise among adult learners (for reviews, see Lin & Zabrusky, 1998; Pieschl, 2009; Stone, 2000) and teachers’ self-reported instructional practices which reveal their underlying beliefs about teaching and learning (e.g., Calderhead, 1996; Richardson, 1996; Swan, 2006).

**Expertise.** Learners’ perceptions of their level of expertise seem to affect their assessments of their learning. Those who perceive themselves to be experts are more likely to highly rate their abilities (Glenberg & Epstein, 1987; Schraw et al., 2013; Stone, 2000). In

contrast, observed expertise leads people to underrate their abilities; as people gain additional knowledge and skills, they become more aware of their own knowledge deficiencies and then perceptions of their ability decrease (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Stone, 2000). Individuals may draw from feelings of confidence related to their performance to gauge their own knowledge and skills, so experiences and perceptions that affect individuals' confidence and self-efficacy are also thought to affect their knowledge estimations (e.g., see Pieschl, 2009; Stone, 2000). Thus, we expected that teachers with perceived expertise in mathematics teaching (e.g., teachers with lots of teaching experience, but not necessarily strong mathematical knowledge for teaching) would provide higher self-assessments of their knowledge and therefore higher self-assessments of their learning (Kruger & Dunning, 1999); and that teachers who have greater skills and expertise in mathematics teaching (e.g., teachers who have strong content and pedagogical content knowledge) would report less learning (e.g., Gigerenzer et al., 1991).

Further, when it comes to the role of expertise in learning captured by direct assessment, several studies have indicated that teaching experience or majoring in a discipline is not associated with gains in teachers' knowledge or skills (e.g., Garet et al., 2016; Garet et al., 2008; Jayanthi et al., 2017). It should also be noted that, even though many studies on teacher learning from professional development measure teachers' background characteristics (e.g., teaching experience and major) few report whether these characteristics are associated with outcomes.

**Self-reported teaching practices.** Other factors that may differently influence teachers' self-reported learning and knowledge gains are their self-reports about classroom practices which are thought to arise from their beliefs about teaching and learning (e.g., Calderhead, 1996; Swan, 2006). A common distinction made in educational research refers to student-centered teaching versus teacher-centered teaching (e.g., Kember & Gow, 1994; Weimer, 2002), which originates

from differing views of the centrality of the roles of teachers and students in the classroom. Briefly, a teacher-centered orientation holds that knowledge is transferred to students through the teacher, the sole deliverer of knowledge in the classroom, whereas a student-centered orientation is the view that students actively construct knowledge through social interaction. We have not found any prior work on the influence of student- versus teacher-centered beliefs and practices on *learning* from professional development; however, De Vries and colleagues (2014) found that among 260 Dutch secondary school teachers, those who reported more student-centered beliefs and practices also reported higher *participation* in professional development activities. We expected, based on this research, that teachers who reported using such practices would report making greater learning gains from professional development because of their perceived greater participation in professional development activities.

### **Present Study**

On the basis of earlier work, we expected that teachers' perceived learning would not be closely aligned with their assessed learning and that it might be associated with different sets of teacher characteristics and self-reported teaching practices. In this study, we used data collected from 545 teachers who participated in yearlong professional development activities and whose learning was measured by both self-reported and direct measures of learning gains. We aimed to answer the following questions:

- To what extent will teachers' self-reports of their gains in MKT align with the MKT gains measured by direct assessments?
- To what extent will teachers' background characteristics and reported instructional practices be associated with their self-assessed and directly assessed MKT gains?

### **Method**

## Context

To conduct this study, we partnered with a professional development organization supported by a Mathematics and Science Partnership grant from the U.S. Department of Education. This organization is part of a statewide network of partnerships that provide professional development to K-12 teachers in mathematics and science. Each year, a number of projects are funded to provide their participants with yearlong content-focused professional development. These institutes are required to create opportunities for the participating teachers to enhance their content knowledge for teaching in the domain of mathematics.

We obtained teachers' background, PD participation, and MKT assessment data of participants from the professional development organization. In addition, this organization distributed the survey we developed to all teachers who were attending the professional development program in mathematics. Specifically, during the time of the study, 24 projects were providing professional development in mathematics to K-12 teachers. These projects began with a summer institute lasting 4 to 10 days, depending on the project, and continued throughout the year with follow-up activities. Teachers in these projects completed various activities designed to enhance their mathematical knowledge and pedagogical skills, totaling on average 114.6 professional development hours ( $SD = 18.3$ ). The facilitators of these projects had on average 7.6 years of professional development facilitator experience ( $SD = 5.5$ ) and 14.5 years of teaching experience ( $SD = 7.6$ ), and 29% of them had PhDs in mathematics or mathematics education. These projects had on average 26.4 teachers ( $SD = 11.1$ ).

## Sample

Our target sample included all the teachers who attended professional development in mathematics provided by these 24 projects in the 2015–2016 academic year. Our analytic sample

included those who had completed direct assessments at both the beginning and end of the professional development program and had completed self-reports on their learning from the same professional development.

Of the 634 teachers who had participated in the professional development programs, 545 had completed our self-report survey, for a response rate of 86%. We checked whether the analytic sample was representative of the full sample and found no statistically significant differences between the teachers who had completed the self-report survey and those who had not based on years of teaching experience ( $M = -.21$ ,  $SD = .91$ ,  $t(624) = -.23$ ,  $p = .82$ ), years participating in professional development programs ( $M = .02$ ,  $SD = .21$ ,  $t(624) = .08$ ,  $p = .94$ ), and educational degree (master's degree or higher;  $\chi^2(1, N = 634) = .0001$ ,  $p = .99$ ).

Almost three-fourths of the teachers in the analytic sample were White (73%), whereas 15.4% were Hispanic and 6.8% were African American. Twenty-nine percent held a master's degree or higher, and their years of teaching experience ranged from 1 to 43 years, with a mean of 10.8 ( $SD = 8.0$ ) and a median of 9. Forty-six percent of these teachers were elementary school teachers, 36% were teaching at the middle school level, and 18% were high school teachers.

[Insert Table 1 around here]

### **Data Collection Procedure**

As part of the grant requirements, each participating teacher was required to take an assessment before the program commenced and after it ended. These assessments were delivered online by the professional development organization to ensure the validity and quality of the data collected. All the teachers in these projects had completed the MKT assessments developed by Hill, Ball, and colleagues (Hill et al., 2004). On average, teachers who participated in these programs had a moderate and statistically significant increase in their MKT from pretest to

posttest, based on the direct assessments ( $M_{change} = .27$ ,  $SD = .73$ ,  $p < .001$ , effect size of .37).

We retrieved data on teachers' pre- and posttest MKT scores, the dates when they completed these assignments, and background information on the teachers from the professional development organization.

In addition to retrieving data from the professional development organization, we collected data from the participating teachers in these programs through an online survey we had adapted from an existing survey used in earlier large-scale studies (e.g., Garet et al., 2001; Desimone et al., 2002). Specifically, teachers were asked to evaluate the learning that had occurred as a result of their participation in the program by rating the extent to which they felt their knowledge and skills had increased in several areas. To ensure that the kinds of knowledge teachers reported gaining were aligned with the knowledge captured by the direct assessments, we utilized validity studies conducted by the MKT instrument developers that identified what kinds of knowledge teachers drew on when they answered these questions (e.g., Hill et al., 2007). These studies indicated that teachers used their understanding of mathematical concepts and the knowledge of their students' mathematical thinking when they answered the items on these assessments; therefore, in our survey, we specifically asked teachers to rate how they felt their understanding of mathematical concepts had deepened, how their understanding of how children think and learn about mathematics had increased, and how their attention to children's thinking and learning when planning their mathematics lessons had increased.

We also captured teachers' reported instructional practices through an existing measure (Swan, 2006). The survey also included a linking variable that would allow us to connect the survey data to the data on teachers' background and direct assessment. Both the direct assessment and the survey we developed was administered through the professional development

organization so that the participants could express their learning freely, knowing that the partner projects would not have access to their self-reports. The survey was sent out at the end of the projects along with the MKT assessments (i.e., posttests) so that individual participant teachers' self-reports and their MKT assessed by the direct measures would be based on the same sets of activities they had completed. Ninety percent of the teachers completed our survey and the posttest on their MKT within 24 hours. Furthermore, we checked the teachers' activity logs to ensure that the rest of the teachers had not completed any additional activities between of the time they completed the content knowledge assessment and the survey. Therefore, we are confident that teachers' self-reports and the direct assessment of their MKT gains were based on the same set of learning experiences.

## Measures

### Outcome measures.

***Direct assessment of teachers' MKT.*** The change in participant teachers' MKT was measured by instruments developed by a team led by Hill and Ball to measure the mathematical knowledge that teachers need in teaching (Hill et al., 2004). The content validity of these instruments was previously established through interviews conducted with elementary school teachers, nonteachers, and mathematicians, and the construct validity was established through factor analysis methods (Hill et al., 2007). After creating assessments for elementary school teachers, they developed assessments for middle school teachers in several mathematical domains, such as proportional reasoning, as well as elementary school versions of the assessments in several content domains (LMT Project, n.d.). Hill (2007) contended that the MKT assessments for middle school teachers could be considered valid measures because the items were created by using "the same construct map, largely the same set of item writers, and the

same item formats and style" (p. 100). Furthermore, some of the newly developed middle school items were administered to a sample of nationally representative middle school teachers in the United States (Hill, 2007).

As mentioned in prior work (Author, 2018), the domains of the MKT framework assessed were narrower than those mentioned in the MKT theoretical framework. Specifically, as illustrated in Figure 1, the majority of the MKT assessments captured teachers' understanding of the mathematics content and students' mathematical thinking.<sup>2</sup> Furthermore, the think-out-loud interviews the MKT instrument developers conducted with teachers indicated that the teachers generally correctly answered the MKT items by using their understanding of mathematics concepts or their attention to students' thinking (Hill et al., 2007). In sum, the instrument developers created assessments that captured similar aspects of MKT in different mathematical content areas, such as patterns, functions, and algebra, for elementary or middle school teachers.

[Insert Figure 1 here]

Thus, depending on the content targeted in a project, the project team used the MKT assessment forms in a corresponding content area to accurately assess the change in teachers' MKT. For instance, if the project targeted the proportional reasoning concept, the proportional reasoning forms were administered to participating teachers to assess the change in their MKT. As shown in Table 1, five different forms ranging from 27 items to 33 items were used across these projects to capture the teachers' MKT according to the content focus of the programs. Table 2 shows that Cronbach's alpha ranged from .81 to .90, suggesting that the forms used in

---

<sup>2</sup> Because existing MKT assessment items are not publicly available, we have included the three sample items shown here, which were published by the MKT instrument developers to illustrate the types of knowledge captured in the assessments used in this study.

the study had internal consistency. Participating teachers' MKT was measured by using parallel forms developed by the MKT instrument developers for each content area, one at the beginning of their professional development activities and the other at the end of their last professional development session (Table 2).<sup>3</sup> Gain scores were created by standardizing teachers' posttest scores based on the pretest mean and standard deviation for each assessment.

[Insert Table 2 around here]

**Scale used for teachers' self-reported MKT.** Because we wanted to ensure that the kinds of knowledge gains measured by these two outcome measures were as similar as possible, we asked teachers to report the kinds of knowledge they drew on when they were answering the MKT items. Specifically, according to the validation studies conducted by the MKT developers (Hill et al., 2007), teachers tapped into their knowledge of mathematics and their knowledge of their students' thinking when they answered the MKT items. Therefore, we measured teachers' self-reported gains in MKT by asking them about improvements in their (1) understanding of mathematics, (2) understanding of students' thinking, and (3) attention to students' mathematical thinking while planning their math lessons. Items were based on a 5-point scale, ranging from 1 (*almost never*) to 5 (*almost always*).<sup>4</sup> The Cronbach's alpha for this scale was 0.87. To obtain the second outcome measure score, we computed averages of the scores on these three items.

---

<sup>3</sup>Hill and Ball developed no MKT assessments for high school mathematics teachers, yet a considerable portion of the teachers in the study were teaching at the high school level. The majority of high school teachers who attended this project were ninth-grade teachers who attended professional development along with other middle school teachers. Nevertheless, when we reran the analyses and excluded these teachers, we found that the results still held true (i.e., the results and the size of the observed relationships were almost identical; see Appendix A).

<sup>4</sup> We asked teachers to think about their professional development experience provided by their specific program during the academic year as their reference and answer the following question: "To what extent do you feel that your knowledge and skills have been enhanced in each of the following areas as a result of your participation in the [program name]?" The following options were given: "Deepening knowledge

**Control variables.**

**Teachers self-reported instructional practices.** We measured teachers' self-reported mathematics teaching practices by using an instrument developed by Swan (2006) as a covariate in our final model. This instrument consists of 25 items that capture the frequency of a specific self-reported classroom behavior on a 5-point scale ranging from 1 (*none of the time*) to 5 (*all the time*). The construct validity of the instrument was previously established (1) by comparing the descriptions of teachers' instructional practices with the items presented in this instrument, (2) by showing that the ratings were consistent with classroom observations, and (3) by showing high correlations between teachers' reported practices and their students' descriptions of their teachers' practices (Swan, 2006, 2007). The Cronbach's alpha reliability coefficient was previously found to be .85 (Swan, 2016). Thirteen of the items on the instrument were designed to capture teacher-centered practices, such as the teacher mainly using whole-class discussion, following the planned materials very closely, avoiding having students make mistakes by first explaining the concepts, teaching each topic from the beginning and assuming that students do not know much, and showing only one way of solving a problem. In contrast, the student-centered teaching practice items were designed to capture how often teachers formed links between mathematical concepts, encouraged students to make and discuss their mistakes, adjusted their teaching based on what the students already knew, and allowed students to invent their own methods and compare different methods. The internal consistency values of the teacher-centered and student-centered scales (Cronbach's alphas) for this study were .79 and .77,

---

of mathematics"; "Understanding of how children think and learn about math"; and "Attention to children's thinking and learning when planning math lessons."

respectively. We obtained teachers' scores on these two scales by averaging their scores on the items in the corresponding scales.

***Teacher characteristics.*** We included many teacher characteristics as covariates in our final model. Dummy variables were included in our analysis to capture teachers' ethnicity (White as the reference category),<sup>5</sup> the grade band in which they taught (elementary [reference category], middle school, or high school), whether they had majored in mathematics or science during their undergraduate education, and whether they had a master's degree or higher. The analysis also included teachers' categorized years of teaching experience (1 indicated 3 or fewer years of teaching experience, 2 indicated 4 to 6 years of teaching experience, 3 indicated 7 to 10 years of teaching experience, 4 indicated 11 to 15 years of teaching experience, and 5 indicated 16 or more years of teaching experience), a continuous variable for the number of years teachers had participated in professional development programs, and their standardized pretest scores on the direct assessment.

### **Analytic Approach**

To investigate how the gains reported by teachers and those measured by direct assessments were related, we first checked the correlations among teachers' self-reports for changes in their MKT, the gains measured by the direct assessment, and the initial mathematical knowledge level measured by the direct assessment. To account for teachers from different projects, we then used two-level hierarchical linear models (teacher level and project level) to investigate which teacher background characteristics predicted the change in teachers' knowledge depending on the measure (self-reports or the direct assessment). An analysis of

---

<sup>5</sup> We included a binary-coded ethnicity indicator because, in prior work, differences were found in the self-reports of White and non-White respondents (e.g., Fendrich & Johnson, 2005; Johnson & Bowman, 2003). Therefore, we added this variable to capture potential factors affecting self-reports.

intraclass correlations justified the use of a two-level model because the projects accounted for, respectively, 5.4% and 9.4% of the variation in teachers' self-reports and in their gains measured by the direct assessment.

We then conducted a separate analysis for the two outcome measures that included the same predictors. Specifically, a teacher's score on the outcome measure (*Gain*) was a function of the teacher's ethnicity, the grade level taught (middle school or high school), a dummy-level indicator of whether the teacher had a master's degree or higher, a dummy-level indicator of majoring in mathematics or science, the teacher's standardized pretest score, her or his years of teaching experience, the number of years the teacher had attended professional development programs, and teachers' self-reported teacher-centered teaching and student-centered instructional practices. All scores were centered around the grand mean. In addition, we included indicator variables for each version of the test to adjust for test-specific differences in the MKT scores. No project-level predictors were included in the data analysis. We used the *xtmixed* command in STATA 14 to estimate the multilevel mixed-effects linear regression models:

Level 1 (teacher):

$$\begin{aligned}
 Gain_{ip} = & \beta_{0p} + \beta_{1p}White_{ip} + \beta_{2p}middle\ school_{ip} + \beta_{3p}high\ school_{ip} \\
 & + \beta_{4p}graduate\ degree_{ip} + \beta_{5p}major_{5p} + \beta_{6p}pretest_{ip} \\
 & + \beta_{7p}teaching\ experience_{ip} + \beta_{8p}PD\ experience_{ip} \\
 & + \beta_{9p}teacher\_centered_{ip} + \beta_{10p}student\_centered_{ip} + \sum test \\
 & + r_{ip}, \quad r_{ip} \sim N(0, \sigma^2),
 \end{aligned}$$

Level 2 (project):

$$\beta_{0p} = \gamma_{00} + u_{0p}, \quad u_{0p} \sim N(0, \sigma_{00}^2), \quad \text{where } \beta_{ip} = \gamma_{ip}, \quad i = 1, 2, \dots, 10.$$

## Findings

### Research Question 1: Relationships Between Perceived and Observed Knowledge Gains

The correlation between teachers' self-reported gains in MKT and their gain scores based on the direct assessment capturing the same construct was almost zero ( $r = -0.0003, p = .99$ ), indicating the underlying constructs assessed by the two measures were different. Additionally, teachers' initial MKT (captured by the direct assessment) was negatively associated with the learning captured by these two measures at different magnitudes. Teachers' scores on the initial test had a weak but significant correlation with their self-reported MKT gains ( $r = -0.09, p = 0.04$ ), whereas initial MKT scores were moderately correlated with their gains measured by the direct assessment ( $r = -0.39, p < .0001$ ).

### **Research Question 2: Predictors of Perceived and Observed Learning**

Table 3 summarizes the results from the two-level hierarchical linear regression models estimating teachers' self-reported gains and their gains measured by a direct instrument as a function of their educational background; and self-reported teaching practices.<sup>6</sup> When the outcome measure was teachers' self-assessment of their learning, four variables had a statistically significant association with teachers' self-reported knowledge gains. Specifically, teachers who majored in mathematics or science reported .12 points less knowledge gain ( $p = .041$ , effect size =  $-.210$ )<sup>7</sup> compared with those whose majors were not mathematics or science. Non-White teachers' reported knowledge gain was .17 points more than that of teachers from White teachers ( $p = .005$ , effect size =  $.307$ ). Teaching experience was positively related to their

---

<sup>6</sup> We ran a robust regression analysis to ensure that the results were not affected by outliers or influential observations. Findings were similar with respect to significance and effect size.

<sup>7</sup> We estimated the effect size by using the formula for effect sizes for predictors in multilevel modeling suggested by Tymms (2004). For dummy variables, we used  $\frac{\beta_1}{\sigma_{teacher-level\ variance}}$ , where  $\beta_1$  is a coefficient for a dummy-coded variable and  $\sigma$  is a Level 1 residual standard deviation. For continuous variables, we used  $\frac{\beta_2 * 2SD_{variable}}{\sigma_{teacher-level\ variance}}$ , where  $\beta_2$  is a coefficient for a continuous variable,  $SD$  is the standard deviation of that variable, and  $\sigma$  is a Level 1 (teacher-level) residual standard deviation.

reported knowledge; teachers in greater teaching experience categories reported .05 larger knowledge gains ( $p = .014$ , effect size = .236). Teachers' self-reported student-centered instructional practices also positively predicted the self-reported gains in MKT ( $b = .28$ ,  $p < .001$ , effect size = .515).

A set of teacher-level predictors was also significantly linked to the knowledge gains measured by the direct assessments. Specifically, teachers who had a master's degree or higher increased their MKT by .19 points more than did those who did not have a master's degree ( $p = .005$ , effect size = .279). The grade level teachers teach positively predicted gains in their MKT (effect sizes of .666 [ $b = .46$ ,  $p < .001$ ] and 1.061 [ $b = .74$ ,  $p < .001$ ] for middle and high school teachers compared with elementary teachers, respectively). Teachers' initial pretest scores on the direct assessment were negatively related to their gain scores, with an effect size of  $-1.259$  ( $b = -.44$ ,  $p < .001$ ). Pretest scores may be negatively associated with learning gains because individuals who know more on the pretest may be able to learn less and regress to the mean. Teachers' scores on the teacher-centered instructional practices were negatively linked to their gain scores, with an effect size of  $-2.79$  ( $b = -.18$ ,  $p = .002$ ).

[Insert Table 3 around here]

## Discussion

The aim of this study was to compare the improvements in MKT reported by teachers and measured by direct assessments. An analysis of data collected from hundreds of teachers who participated in different professional development programs indicated no correlation between teachers' self-reports and the direct assessments of their learning. This finding is especially important given that we explicitly tried to increase the congruence in what was captured by these two outcome measures by asking teachers to report changes in the aspects of

knowledge they drew on when answering items on the direct assessments. Furthermore, our findings suggest that different sets of teacher background characteristics and self-reported instructional practices are associated with the learning captured by these two outcome measures.

We first discuss what the lack of correlation between self-reported and directly assessed learning means for professional development design and research. This result suggests that teachers' self-reports and the direct assessments captured different underlying constructs. Therefore, a program that is identified as effective based on teachers' self-reports might not be considered effective if the outcome measure was a direct assessment of teachers' learning. Note that several large-scale studies that have played a key role in determining what makes professional development effective are based on teachers' self-reports, (e.g., Garet et al., 2001; Desimone et al., 2002). Thus, our findings suggest caution is warranted when evaluating conclusions drawn from these studies. Our findings also urge researchers and teacher educators to explicitly pay attention to the measures used to evaluate the success of professional development programs. Focusing on what is captured by the outcome measures used to evaluate the success of a program can influence professional development leaders to identify what learning opportunities need to be revised. We also contend that multiple measures should be used to capture different aspects of teachers' learning, to accurately depict the impact of a program on teachers, and to better understand the interactions among different aspects of teachers' learning that are targeted in a program.

Another important facet of this finding is related to teachers' uptake of the professional development initiatives. In a meta-analysis of adult learning and workplace training studies, Sitzmann et al. (2010) found self-assessed learning and self-assessed knowledge to be more strongly related to affective constructs—such as learners' satisfaction with their instructional

experiences, their motivation to apply the skills learned in a program, and their confidence in their ability to perform the newly learned tasks—than with direct assessments. Thus, teachers' self-assessments may be an indicator of how confident they feel in their ability to apply what they have learned rather than how much they have actually learned from the program. Moreover, teachers' confidence in their ability to perform a task is linked to their willingness to adopt innovative practices (e.g., Smylie, 1988). Thus, teachers' confidence may determine how they interpret and act upon their learning experience. For example, consider a teacher who did not feel like they learned much from a program despite their gains on direct assessments. This teacher might not be willing to implement what they have learned from professional development in their teaching practices even though their knowledge of how to perform these tasks might have increased according to the direct assessments. Furthermore, such a teacher might needlessly devote their time and attention to relearning knowledge or skills they have already gained. Alternatively, consider a teacher whose knowledge did not increase substantially according to a direct assessment but felt as if their knowledge and skills had changed. This teacher might experience a confidence boost, which in turn might influence their attitudes toward implementing the new knowledge and skills they perceived as gaining from the program (Guskey, 1988). However, these teachers who gained confidence in their new skills, but not the skills themselves, might not be able to implement the new skills or knowledge effectively. With these examples in mind, it becomes clear that different combinations of perceived and observed learning from a professional development program may lead to different consequences for teacher change.

Our findings also suggest that different teacher-related factors were associated with the learning reported by self-assessment versus direct assessment. Teachers' ethnicity,

undergraduate preparation, and teaching experience were associated with their perceptions of learning, whereas the grade band they taught and the highest level of education they had achieved were associated with the learning gains captured by the direct assessment. These results are consistent with prior literature in that the perceived level of expertise seemed to influence teachers' assessment of their learning (e.g., Glenberg & Epstein, 1987; Schraw et al., 2013; Stone, 2000). Specifically, teachers who majored in mathematics and science in their undergraduate education reported learning less could be because of their higher self-assessment of their baseline MKT. Similarly, teachers with observed expertise (i.e., those with more years of teaching experience) reported learning more again possibly because of a greater awareness of their own knowledge deficiencies (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Stone, 2000). However, teachers' perceived expertise was not related to their observed learning.

It is interesting that teachers' self-reported instructional practices were related to their perceived and observed learning. Teachers who reported employing instructional practices to promote student-centered learning seemed to feel that they learned more from the program even though their learning did not differ according to the direct assessments. These findings are consistent with studies that found associations between student- and teacher-centered beliefs with professional engagement. Prior research found that teachers' student-centered beliefs correlated with greater reported engagement in collaborative professional development activities, whereas more traditional, teacher-centered beliefs were negatively correlated with professional engagement (Becker & Riel, 2000; De Vries et al., 2015). Teachers with more student-centered dispositions may have been more engaged in professional development activities and consequently felt as if they learned more. Another explanation is that, because the practices in the professional development programs investigated in this study resembled student-centered

teaching practices, teachers who reported using student-centered practices might have assumed that they learned more because of the similarities between what they saw in professional development and what they felt they did in their teaching. As such, teachers who espoused student-centered teaching practices may have had a more favorable learning experience, leading to increased confidence and self-assessed learning, while learning about the same as their peers. In contrast, the more teachers reported using teacher-centered practices, the less they seemed to learn from professional development. This could be related to the fact that these teachers may have been less engaged in the program because of a lack of alignment between their teaching practices (and therefore their underlying beliefs about teaching and learning) and the teaching practices used in the professional development program. This in turn could have created fewer learning opportunities for these teachers. It is interesting to note that, although these teachers did not feel they learned less according to their self-reports, the direct assessments indicated that they had learned less.

### **Limitations**

Before discussing the implications of the findings, we would like to note the limitations of the study. First, compared with K-12 mathematics teachers in the United States, our sample included relatively fewer White teachers as well as more female teachers; hence, our findings should be interpreted with caution. Second, the direct assessments utilized in this study were not designed for high school mathematics teachers. The high school teachers who participated in the study were mainly ninth-grade teachers who attended professional development on algebra and geometry with the middle school teachers. Because the instruments were originally validated with elementary and middle school teachers, it is possible that the measures may not accurately

capture 9th grade mathematics teachers' MKT. However, the results were similar when high school teachers were excluded from the analysis (see Appendix A).

### **Implications for Policy, Practice, and Research**

Our finding that teachers' self-reported learning was not associated with the learning captured by direct assessments suggests that policymakers and practitioners should carefully consider how the primary outcomes of professional development are defined and measured. Different conclusions might be drawn regarding the effectiveness of a given professional development program depending on the outcome measures of teacher learning used. Therefore, we recommend that teacher educators and policymakers carefully consider the forms of knowledge and skills they are interested in promoting during professional development programs and use measures that capture the targeted knowledge and skills. This will oftentimes mean that teacher educators may want to include multiple measures to capture the various aspects of teacher learning targeted in a program.

Specific attention to the measures utilized to determine teachers' learning is also essential to create a knowledge base for professional development and to systematically improve teachers' learning. Because professional development creators use features of successful programs to design their programs, misidentifying or mistakenly concluding what works and what does not work in professional development could have serious consequences for future program design. This is especially a concern for one of the most agreed-on features of effective professional development: the content focus. Because of the wide range of interpretations regarding what content and pedagogical content knowledge entail (cf. Hill, Ball, & Schilling, 2008; Kersting et al., 2012; Rowland, Huckstep, & Thwaites, 2005; Tchoshanov, 2011) and the scant attention paid to what specific kinds of knowledge are measured (e.g., Garet et al., 2016), little is known

about how to create content-focused programs that are effective in enhancing the kinds of knowledge needed for teaching and student learning.

## **Future Work**

Our study provided evidence of the discrepancies between teachers' perceived and observed learning, yet questions remain about how these results might affect changes in the teachers' practices and their students' learning. Given that we relied on the MKT instrument developers' validity work and did not conduct our own validation for this study context, we believe further work with valid assessments is needed to replicate our study. Future research might also focus on how various combinations of teachers' learning and perceived learning outcomes play out in classroom instruction and their impact on subsequent student learning outcomes. Future work on how teachers' perceptions of their learning and their assessed learning affect their instructional decisions and students' learning would be a useful contribution to the teacher learning literature. Additionally, future studies might consider the direct and indirect effects of teachers' beliefs and reported practices on their learning.

## **Conclusions**

This study has important implications for teacher educators, researchers, and policymakers and encourages researchers in the field to pay careful attention to the measures used to assess the impact of a professional development program on teachers' learning. As we found in our study, teacher learning differs by the outcome measure; therefore, the success of the same program could be evaluated differently depending on the measure. Furthermore, what teachers felt they learned from the professional development program and the learning measured by direct assessments were related to disparate sets of teacher and teaching characteristics.

Hence, to make informed decisions regarding how professional development programs will lead to teacher learning, more research is needed on how various aspects of teacher learning can be improved during the same program and how teachers' background characteristics influence their learning.

## References

Author (2018).

Author (2015).

Ball, D.L., & Hill, H.C. (2008). Mathematical Knowledge for Teaching (MKT) measures. Mathematics released items 2008. Retrieved 12th of September, 2010, from [http://sitemaker.umich.edu/lmt/files/LMT\\_sample\\_items.pdf](http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf)

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). Report of the 2012 National Survey of Science and Mathematics Education. *Horizon Research, Inc.(NJ1)*.

Blank, R. K., de las Alas, N., & Smith, C. (2008). *Does teacher professional development have effects on teaching and learning? Analysis of evaluation findings from programs for mathematics and science teachers in 14 states*. Washington, DC: Council of Chief State School Officers.

Bradburn, N. M. (2000). Temporal representation and event dating. In A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain, V. S. (Eds), *The science of self-report: Implications for research and practice* (pp. 49-61). Malwah, NJ: Erlbau.

Calderhead, J. (1996). Teachers: beliefs and knowledge. In D. C. Berliner & R. C. Calfee (eds.), *Handbook of Educational Psychology* (pp. 709–25). New York: Macmillan.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine*,

*verity and fable in the organizational and social sciences* (pp. 311–338). New York: Routledge.

De Vries, S., Van de Grift, W. J., & Jansen, E. P. (2014). How teachers' beliefs about learning and teaching relate to their continuing professional development. *Teachers and Teaching*, 20(3), 338-357.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, 77(5), 1268-1281.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, 20(4), 374-380.

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 247–296). San Diego, CA: Elsevier.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69-106.

Fendrich, M., & Johnson, T. P. (2005). Race/ethnicity differences in the validity of self-reported drug use: Results from a household survey. *Journal of Urban Health*, 82(3), iii67-iii81.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Report prepared for the Institute of Education Sciences. Washington, DC: Institute of Education Sciences.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE 2016-4010). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., and Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological review*, 98(4), 506.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15(1), 84-93.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 4(1), 63-69.

Heck, D. J., Banilower, E. R., Weiss, I. R., & Rosenberg, S. L. (2008). Studying the effects of professional development: The case of the NSF's local systemic change through teacher enhancement initiative. *Journal for Research in Mathematics Education*, 113-152.

Hill, H. C. (2007). Mathematical Knowledge of Middle School Teachers: Implications for the No Child Left Behind Policy Initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95–114.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400.

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, 5 (2-3), 81-92.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, 105(1), 11-30.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives*, 13(10).

Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). *Impact of the developing mathematical ideas professional development program on grade 4 students' and teachers' understanding of fractions* (REL 2017-256). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

Johnson, T. P., & Bowman, P. J. (2003). Cross-cultural sources of measurement error in substance use surveys. *Substance Use & Misuse*, 38(10), 1447-1490.

Kaufman, J. H., Stein, M. K., & Junker, B. (2016). Factors associated with alignment between teacher survey reports and classroom observation ratings of mathematics instruction. *The Elementary School Journal*, 116(3), 339-364.

Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *The Journal of Higher Education*, 65(1), 58-74.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568-589.

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.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121.

Learning Mathematics for Teaching (n.d.). Content and Tasks Measured. Retrieved June 24, 2019 from: <http://www.umich.edu/~lmtweb/content.html>

Lin, L. M., & Zabrocky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, 23(4), 345-391.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29-45.

Newfield, J. (1980). Accuracy of teacher reports: Reports and observations of specific classroom behaviors. *The Journal of Educational Research*, 74(2), 78-82.

Ottmar, E. R., Rimm-Kaufman, S. E., Larsen, R. A., & Berry, R. Q. (2015). Mathematical knowledge for teaching, standards-based mathematics teaching practices, and student achievement in the context of the responsive classroom approach. *American Educational Research Journal*, 52(4), 787-821.

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning*, 4(1), 3-31.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. L., & Schneider, S. A. (1993). *Reform up close: A classroom analysis*. Madison, WI: Wisconsin Center for Research in Education.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (pp. 102–119). New York: Simon & Schuster Macmillan

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, 6(1), 43-74.

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education*, 34(4), 344-363.

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8(3), 255–281.

Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2011). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness*, 4(1), 1–24.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition Learning*, 4, 33-45.

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, 24, 48-57.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure?. *Academy of Management Learning & Education*, 9(2), 169-191.

Smylie, M. A. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal*, 25(1), 1-30.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Erlbaum.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12(4), 437-475.

Swan, M. (2006). Describing mathematics teachers' beliefs and practices: Resources for evaluating design interventions. *Research in Education*, 75(1) 58-70.

Swan, M. (2007). The impact of task-based professional development on teachers' practices and beliefs: A design research study. *Journal of Mathematics Teacher Education*, 10(4-6), 217-237.

Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics*, 76(2), 141-164.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco, CA: Josey-Bass.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues and Answers Report, REL 2007 No. 033). Washington, DC: U.S. Department of Education, Regional Educational Laboratory Southwest.

Zwart, R. C., Wubbels, T., Bergen, T., & Bolhuis, S. I(2009). Which characteristics of a reciprocal peer coaching context affect teacher learning as perceived by teachers and their students?. *Journal of Teacher Education, 60*(3), 243-257.

Table 1

*Descriptive Statistics for Teacher-Level Variables*

Variable	<i>M</i>	<i>SD</i>	%
Teacher background			
Gender (Female)			86.8
Ethnicity (White)			73.8
Educational levels (master's or higher degree)			29.2
Math or science major			30.5
Teaching level			
Elementary school			46.2
Middle school			35.8
High school			18.0
Teaching experience	10.81	7.95	
Years of attending PD programs	1.89	1.11	

*Note.* PD = Professional Development

Table 2

*MKT Instruments Used in the Study*

Instrument	Form	Total items	Reliability
Elementary LMT: Patterns, Functions, and Algebra	A	29	.84
	B	27	.85
Elementary LMT: Numbers and Operations	A	28	.81
	B	29	.83
Middle School LMT: Patterns, Functions, and Algebra	A	33	.89
	B	33	.90
Middle School LMT: Proportional Reasoning	A	30	.86
	B	30	.87
Grades 4–8 Geometry Algebra	A	31	.85
	B	30	.88

*Note.* MKT = Mathematical Knowledge for Teaching; LMT = Learning Mathematics for Teaching Project (University of Michigan).

Table 3  
*Self-Reported Gains and Gains as Measured by the Valid Instrument*

	Self-Reported Gains		Gains Measured by a Direct Assessment	
	<i>b</i> ( <i>SE b</i> )	<i>Effect Size</i>	<i>b</i> ( <i>SE b</i> )	<i>Effect Size</i>
<b>Background Characteristics</b>				
Math or Science Major (Yes)	<b>-.119*</b> (.058)	<b>-0.210</b>	.024 (.072)	0.035
Masters' degree or higher (Yes)	-.049 (.056)	-0.087	<b>.195**</b> (.069)	<b>0.279</b>
Non-White Teachers	<b>.174**</b> (.063)	<b>0.307</b>	-.143 (.076)	-0.205
Middle School Teachers	-.059 (.072)	-0.103	<b>.464***</b> (.089)	<b>0.666</b>
High School Teachers	-.027 (.095)	-0.048	<b>.740***</b> (.116)	<b>1.061</b>
Teaching Experience (0-3yrs=0; 4-6yrs=1; 7-10yrs=2; more than 10-15 yrs=3; more than 15 years=4)	<b>.046*</b> (.019)	<b>0.236</b>	-.009 (.023)	-0.039
Years in the Project	.036 (.024)	0.142	-.036 (.030)	-0.116
Prior Knowledge on MKT	-.004 (.028)	-0.014	<b>-.438***</b> (.034)	<b>-1.259</b>
<b>Self-reported Instructional Practices</b>				
Student-centered teaching	<b>.279***</b> (.050)	<b>0.515</b>	-.102 (.061)	-0.154
Teacher-centered teaching	-.053 (.048)	-0.099	<b>-.181**</b> (.059)	<b>-0.279</b>

*Note.* \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

Students sometimes remember only part of a rule. They might say, for instance, “two negatives make a positive.” For each operation listed, decide whether the statement “two negatives make a positive” statement works, always works, or never works. (Mark **SOMETIMES**, **ALWAYS**, **NEVER**, or **I'M NOT SURE**)

	Sometimes works	Always works	Never works	I'm not sure
a. Addition	1	2	3	4
b. Subtraction	1	2	3	4
c. Multiplication	1	2	3	4
d. Division	1	2	3	4

Students in Mr. Hayes' class have been working on putting decimals in order. Three students--Andy, Clara, and Keisha—presented 1.1, 12, 48, 102, 31.3, .676 as decimals ordered from least to greatest. What error these students making? (Mark **ONE** answer.)

- a. They are ignoring place value.
- b. They are ignoring the decimal point.
- c. They are guessing.
- d. They have forgotten their numbers between 0 and 1.
- e. They are making all of the above errors.

Ms. Abdul is preparing a unit to introduce her students to proportional reasoning. She is considering three versions of a problem that are the same except for the numbers used. Which version of the Mr. Short and Mr. Tall problem below is likely to be the most challenging for students? (Circle **ONE** answer.)

- a. A picture depicts Mr. Short's height as 4 paper clips and as 6 buttons. The height of Mr. Tall (not shown) is given as 6 paper clips. How many buttons in height is Mr. Tall?
- b. A picture depicts Mr. Short's height as 4 paper clips and as 7 buttons. The height of Mr. Tall (not shown) is given as 5 paper clips. How many buttons in height is Mr. Tall?
- c. A picture depicts Mr. Short's height as 2 paper clips and as 9 buttons. The height of Mr. Tall (not shown) is given as 5 paper clips. How many buttons in height is Mr. Tall?
- d. All three of the problems are equally challenging.

*Figure 1.* Sample elementary and middle school mathematical knowledge for teaching (MKT) items (Ball & Hill, 2008).

## Appendix A

*Self-Reported Gains and Gains as Measured by the Valid Instrument (High School Teachers Are Excluded)*

	Self-Reported Gains	Gains Measured by a Direct Assessment	
		<i>b</i> ( <i>SE b</i> )	<i>b</i> ( <i>SE b</i> )
<b>Background Characteristics</b>			
Math or Science Major (Yes)	<b>-.148*</b> <b>(.064)</b>	-.012 (.083)	
Masters' degree or higher (Yes)	-.088 (.061)	<b>.208**</b> <b>(.079)</b>	
Non-White Teachers	<b>.141*</b> <b>(.061)</b>	-.110 (.084)	
Middle School Teachers	<b>-.143*</b> <b>(.057)</b>	<b>.364***</b> <b>(.086)</b>	
Teaching Experience (0-3yrs=0; 4-6yrs=1; 7-10yrs=2; more than 10-15 yrs=3; more than 15 years=4)	<b>.049*</b> <b>(.020)</b>	-.032 (.026)	
Years in the Project	.037 (.026)	-.057 (.036)	
Prior Knowledge on MKT	-.005 (.029)	<b>-.426***</b> <b>(.040)</b>	
<b>Self-reported Instructional Practices</b>			
Student-centered teaching	<b>.265***</b> <b>(.052)</b>	-.121 (.068)	
Teacher-centered teaching	-.082 (.052)	<b>-.288***</b> <b>(.068)</b>	

*Note.*  $N = 447$ . \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$