Copyright © 2015 by the National Council of Teachers of Mathematics, Inc., www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in other formats without written permission from NCTM.

Developing a Mathematics Instructional Practice Survey: Considerations and Evidence*

Michele B. Carney
Boise State University

Jonathan L. Brendefur
Boise State University

Gwyneth R. Hughes
Boise State University

Keith Thiede
Boise State University

As mathematics teacher educators, it is imperative that we have high-quality tools that conceptualize and operationalize mathematics instruction for large-scale examination. We first describe existing instructional practice survey scales, including their conceptualization of practice and related validity evidence. We then present the framework and initial validity evidence for our mathematics instructional practice survey. Survey participants were in-service teachers in a statewide mandated mathematics professional development course. Statistical analyses indicate the items measure two constructs: social-constructivist and transmission-based instructional practice. Of particular interest is the result that these two constructs were negligibly correlated. This is in contrast to the generally accepted notion that social-constructivist and transmission-based instructional practices are the two polar ends of a single construct for describing instructional practice.

Key words: Instructional Practice; Mathematics; Survey Development; Validation

Over the past few decades, educational practice and research have shifted to be more data driven, as is evidenced by both public policy (NCLB, Race to the Top, Common Core) and research initiatives (Institute of Education Sciences & National Science Foundation, 2013). Teachers' instructional practices in the classroom have been shown to be central to student achievement (Boaler, 2005; D'Agostino, 2000), yet studying these practices objectively is inherently difficult given the complexity of classroom instruction. Despite these difficulties, there is a need to define and measure instructional practice in order to evaluate programs and conduct quantitative educational research. But how does one decide what "instructional practice" is, and how does one measure it?

To examine practice meaningfully it must be explicitly conceptualized and then translated into an instrument that appropriately operationalizes and measures this conceptualization. Too often one of two things occurs: Instruments are developed without clear conceptualization of practice, or instruments are used in research or evaluation that do not adequately match the conceptualization of practice within the project. Either disconnect can lead to a lack of significant findings due to the measure rather than the variables under examination. The mathematics education community needs multiple measures of instructional practice that vary in conceptualization and that can be used on a large scale (e.g., self-report surveys) to examine and evaluate different aspects of instructional practice.

The goals of this article are twofold. We first examine existing self-report measures of instructional practice, discussing the conceptualization of instructional practice within each and associated validity evidence. We then conceptualize instructional practice as it relates to our own professional development model and describe why and how we developed our own instrument to operationalize those practices. Lastly, we present the results of two validity evidence studies on our instrument. This article is meant both as a tool for other researchers in thinking about implementing or developing measures of instructional practice, and a report on our own findings with regard to our specific survey instrument.

Operationalizing Instructional Practice—An Overview

Although self-report surveys are useful for large-scale studies, using them to measure a construct as complex as instructional practice is inherently crude. The relativistic nature of what "good" teaching means is an inherent problem in the reliability of teachers' assessments of the quality of their own teaching (Mayer, 1999; Wubbels,

* This manuscript was edited by Margaret Smith, editor of the *MTE* at the time the manuscript was initially submitted.

1992). However, Mayer (1999) found that teachers' reports of *frequency* of specific instructional practices were relatively accurate as compared with the validity issues that arise when teachers report on the quality of instructional practice. Analysis of practice from a frequency perspective, instead of a quality perspective, may increase the agreement between teachers' self-reported practice and observations of actual classroom practice (i.e., the validity of the measure). But it decreases the specificity of what can be detected by the measure, hence its inherent crudeness. This does not mean we should not attempt to measure instructional practice through a self-report survey. Rather, findings based on the survey data should be interpreted relative to the type of measurement that can accurately be performed and the validity evidence provided to support those findings.

Establishing Validity Evidence for Survey Measures of Instructional Practice

Mathematics teacher educators must be critical consumers of research instruments. To this end, we provide a brief explanation of how one establishes validity evidence for psychometric instruments and an overview of commonly used survey scales of instructional practice. Our purpose in this section is to provide mathematics teacher educators with a framework for evaluating validity evidence for survey scales of instructional practice, including our own.

In research we often need to measure a concept of interest. Some concepts can be directly measured, such as height. However, in educational research we often need to measure human-formulated concepts, termed *constructs*, such as mathematics instructional practice. Such constructs are indirectly measured through instruments, such as surveys, that typically contain a grouping of items that together measure the construct. Following appropriate statistical analyses, the individual item scores are combined to provide a scale score reflecting the degree to which an individual demonstrates or possesses a particular construct. Validity for an instrument refers to the strength of evidence that the scale score accurately represents the level an individual possesses of the construct of interest. For example, teachers' ability to respond to student thinking in the mathematics classroom may be the construct of interest. Survey developers may build a scale around that construct with several items that, when combined into a scale score, reflect the nature of a teacher's ability in that area. However, in addition to providing the scale items, developers should provide evidence demonstrating that a high (or low) scale score on the survey actually reflects the presence (or absence) of the construct of interest. As mathematics teacher educators, our selection of instruments should be informed by our particular research context and question(s), and we need to closely examine the validity evidence provided.

In discussing a particular instrument, researchers often refer to the *validity* of the instrument itself. However, current theory focuses on establishing the validity of the evidence that supports the interpretation of an instrument's scores, specific to the context in which the validity evidence was collected (Kane, 2006). For example, if a survey has been validated by evidence from classroom observations of elementary preservice teachers, then the same validation evidence may not support interpretation of scale scores from high school mathematics teachers. The burden is on the researcher to select instruments with validity evidence supporting use of the instrument within their particular research context. We have found that the validation framework from Cook and Beckman (2006), consisting of content, response process, internal structure, and relationship to other variables (described in the paragraphs that follow), provides a thorough yet simple approach to considering validity evidence for psychometric instruments.[1]

Content. Does the instrument measure the depth and breadth of the construct under examination? This involves first clearly defining how the construct is conceptualized and how it will be operationalized. The developer must then provide evidence that the instrument addresses the full depth and breadth of the construct by, for example, providing a framework for the development of the items. If there is no clear evidence of content validity, it does not mean the developers did not have a clear conceptualization or framework, just that this evidence has not been explicitly provided. In that case the burden shifts to the researcher who plans to use the scale to ensure it addresses the depth and breadth of the construct as needed for their research or evaluation purpose.

Response process. Is there a match between the thought process of the respondent and the intended construct under examination? For surveys, this involves providing evidence that respondents are interpreting the survey items as intended, such as by conducting cognitive interviews (Desimone & Le Floch, 2004). A related consideration when researchers intend to use survey scales to evaluate change from pre- to post-intervention is response-shift bias (Bray, Maxwell, & Howard, 1984), which involves validity threats that arise when respondents' understanding and interpretation of survey items shift as a result of participation in an intervention. For

1 Cook and Beckman (2006) also address consequential validity.

example, teachers' understanding of a survey item related to what it means to ask students to justify their answers can change significantly as a result of professional development. This can often be addressed through a retrospective pre/post survey design that involves asking participants to recall their preintervention status postintervention (Lam & Bengo, 2003).

Internal structure. Is there an acceptable level of evidence regarding internal consistency and construct structure? This refers to evidence that the measurement error around a set of items is minimal (reliability) and that the items collectively measure a common construct (unidimensionality). Cronbach's alpha is often used to measure the internal consistency of a set of survey items. It provides an indicator of the correlation between survey items and the likelihood of obtaining similar responses if readministered. A low Cronbach's alpha indicates the survey items may not be accurately or consistently measuring the construct under investigation and suggests that a researcher should not compute a scale score for the survey items because they are measuring more than one concept of interest. In addition to Cronbach's alpha, the use of statistical techniques, such as factor analysis (FA), examines the unidimensionality of the survey scale, ensuring the items measure a single construct rather than multiple constructs that are correlated with each other.

Relationship to other variables. Is there an appropriate level and direction of correlation to other variables of interest? This involves multiple aspects of validity evidence, including criterion, convergent/discriminate, concurrent, and predictive. For example, criterion validity typically refers to evidence that the instrument under examination produces scores that are correlated to a related criterion measure. For the construct of instructional practice, this could be demonstrated by providing a correlation coefficient between self-report survey scale scores of instructional practice and observations of classroom practice. Predictive validity evidence indicates the ability of one variable (independent) to predict another (dependent). For example, instructional practice is often examined in relation to student achievement measures. However, due to the multitude of factors influencing student achievement, the expected predictive between these variables is often difficult to demonstrate.

As may be evident from the description of the various elements of the validity framework, the process of creating, modifying and examining the validity evidence for survey items and scales is an intensive, iterative process that should not be undertaken lightly. Ideally, researchers will identify a scale in the literature that conceptualizes practice in a manner that meets their needs. Below we provide a brief description of four instructional practice survey measures and their associated validity evidence.

Construct Validity Evidence for Existing Instructional Practice Survey Instruments

There are multiple measures of instructional practice available for researchers and practitioners to use in research and evaluation. Our review focused on identifying self-report instructional practice survey scales with the following characteristics: (1) they are described in published, peer-reviewed articles focused on presenting validity and reliability evidence, (2) the sample used inservice (versus preservice) teachers, and (3) the response categories focused on frequency of occurrence of particular practices (versus level of agreement/beliefs). Due to the limited number of peer-reviewed articles identified, we also included reports of survey development and validation that were not peer reviewed but provided sufficiently detailed analysis. We identified four survey scales that fit these criteria. In the following section we briefly examine these survey scales through the lens of content, response process, internal structure and relationship to other variables (Cook & Beckman, 2006). Table 1 also provides some general information about each of the scales. Please keep in mind our analysis is based on the information we found in the documents we reviewed. Because it is difficult to publish the full breadth of survey validation evidence, it may be that additional research was conducted but not reported or was missed in our review process.

Horizon-Reform. This four-item scale assesses the use of "reform-oriented teaching practices" and is a part of the much larger Horizon Research 2012 National Survey of Science and Mathematics Education (Banilower et al., 2013). The construct of reform-oriented teaching practices is conceptualized as students' use and explanation of multiple approaches to solving a mathematics task. The survey report indicates *response process* was examined through cognitive interviews that were conducted for the entire survey (Banilower et al., 2013, p. 4). *Internal structure* evidence was provided through examination of the unidimensionality of the items in the scale through factor analysis and providing a Cronbach's alpha (Banilower et al., 2013, pp. E–16). Due to the overall breadth of the National Survey of Science and Mathematics Education, there is little content validity evidence provided for this particular scale, and *relationship to other variables* was not addressed in the report.

TIMSS-Engagement. This four- or six-item scale for eighth- or fourth-grade teachers, respectively, assesses the use of "instruction to engage students in learning"

Table 1
General Survey Information on a Sample of Published Self-Report Instructional Practice Survey Measures for Practicing Teachers

Survey	Original purpose	Sample	Constructs (No. of items)	Internal structure	Sample item
Horizon-Reform	Evaluate mathematics instructional practice from a national sample and examine relationship to equity variables	National samples of U.S. practicing elementary, middle, and high school teachers	Reform-oriented teaching practices (4)	α = .77 and CFA	*Have students explain and justify their method for solving a problem.*
TIMMS-Engagement	Provide comparison of mathematics instructional practice across countries and relationship to achievement	International samples of practicing 4th- and 8th-grade teachers	Engage students in learning (4th: 6 items) (8th: 4 items)	4th: α = .4 to .83 8th: α = .18 to .76 and CFA	*Praise students for good effort.*
Swan-Practices	Evaluate the influences of a mathematics professional development project.	184 FE teachers (two groups)	Instructional practices (25 items)[1]	α = .85 No CFA	*I am surprised by ideas that come up in a lesson.*
Ross-Commitment	Evaluate the commitment to implementation of standards-based mathematics teaching	2687 K–8 teachers in U.S. (two groups)	Commitment to standards-based mathematics teaching (20)[1]	α = .81 to .88 No CFA	*I tend to integrate multiple strands of mathematics within a single unit.*

[1] Used reverse coding of particular items (i.e., viewed instructional practice as a continuum).
CFA = Confirmatory factor analysis

and is part of the International TIMSS 2011 Context Questionnaire administered to teachers whose students participate in the Trends in International Mathematics and Science Study (TIMSS) assessment (Martin & Mullis, 2012). "Instruction to engage students in learning" is conceptualized as encouraging, engaging, and questioning students during various aspects of mathematics instruction. *Internal structure* evidence was provided through examination of the unidimensionality of the items within the scale using factor analysis. From an internal-consistency perspective, the scale analyses offer an interesting perspective on the importance of considering the context and sample from which the validity evidence is drawn. The six-item scale had an alpha that ranged from .40 for teacher respondents in Azerbaijan to .83 in Chinese Taipei. The range in the four-item scale was considerably larger, .18 to .76. These large discrepancies in scale reliability across countries emphasize the need to consider the sample and context from which the internal structure evidence is drawn when determining whether or not the scale will hold together in a different research or evaluation context. *Relationship to other variables* was examined through correlational analysis, and essentially no consistent relationship between student responses and students' performance on the TIMSS were found. *Content* and *response process* were not specifically addressed within the report.

Swan-Practices. This 25-item scale assesses "teacher-centered practices" arising from a transmission-oriented belief system or "student-centered practices" arising from a constructivist-oriented one (Swan, 2006). "Teacher-centered practice" is conceptualized as the teacher transmitting knowledge to students, while "student-centered practice" is conceptualized as taking students' individual processes for knowledge-building into account through instruction designed to be flexible to student needs. This questionnaire was developed in conjunction with a beliefs questionnaire to assess changes as a result of professional development provided by Further Education colleges in England. *Content* validity evidence is provided through the framing of practice as arising from transmission, discovery, or constructivist beliefs about mathematics teaching and learning (Ernest, 1989). *Response process*

evidence was provided through the use of instructional practice vignettes, Cronbach's alpha was reported as a mean of evidence for *internal structure*, and observations of classroom practice and students' descriptions of teachers' practices were provided as criterion validity evidence for *relationship to other variables*.

Ross-Commitment. This 20-item scale assesses elementary teachers' "commitment to mathematics education reform" (Ross, McDougall, Hogaboam-Gray, & LeSage, 2003). It conceptualizes this commitment a broad array of topics, from giving students open-ended problems to the teacher acting as a facilitator rather than a transmitter of knowledge. The authors provide *content* validity evidence through a clear framework for the development of the survey items based on the NCTM standards (2000) and examination of items by a panel of elementary mathematics specialists. *Internal structure* evidence is given in the form of Cronbach's alpha, and evidence of *relationship to other variables* is provided through examination of the relationship between the scale score and school-level means for a test of sixth grade mathematics achievement. *Response process* is not specifically addressed in the article.

Operationalizing Instructional Practice—the DMT Framework

We chose to undertake the development and validation process for our instrument because we wanted the operationalization of instruction to closely match the Developing Mathematical Thinking (DMT) framework for instructional practice so we could measure the influence of our professional development on a large scale. The existing measures either did not closely match on conceptualization of practice, or in some cases they conceptualized practice as a continuum (e.g., Swan-Practices and Ross-Commitment), which could be measured by a single scale. We found this problematic based on our experience in classrooms observing teachers' instructional practice. It is important to note we did not undertake the development of our own survey scale lightly. The development of a survey scale and process for gathering and presenting validity evidence is extremely complicated. The decision to develop one's own scale should be considered in light of the time and resources needed to create a high-quality scale with regard to the various validity elements.

For our purposes, the Swan-Practices scales, which use learning theories as a lens for thinking about instructional practice, provided a basis for conceptualizing our scales. We chose to develop our own items rather than use the Swan-Practices scales based on our fundamentally different conceptualization of instructional practice. In particular, we did not want to make the assumption that what are commonly referred to as teacher-centered and student-centered practices lie at opposite ends of a single construct's continuum. We instead view student-centered and teacher-centered practices as related but independent subconstructs within instructional practice. The next sections provide content validity evidence by describing the conceptual domains of the DMT framework and the development of the initial survey framework and items.

Developing Mathematical Thinking (DMT) Theoretical Framework

The DMT framework is built upon social and cognitive learning theories, which hold that students need to learn mathematics by constructing knowledge through meaningful classroom activities and discussions. The teacher's role in the classroom is to facilitate student learning through the meaningful selection of mathematical tasks and high-quality classroom discussion designed to build connections between students' informal knowledge and the formal knowledge of mathematics that has developed over time (Carpenter, Fennema, Franke, Levi, & Empson, 1999; Gravemeijer & van Galen, 2003; Hiebert, 1997). Our framework focuses on five classroom instructional practices that develop students' mathematical understanding: (1) taking students' ideas seriously, (2) pressing students conceptually, (3) encouraging multiple strategies and models, (4) addressing misconceptions, and (5) focusing on the structure of the mathematics (Brendefur, Carney, Hughes, & Strother, forthcoming; Carney, Brendefur, Thiede, Hughes, & Sutton, 2014). A brief description of each of the domains follows.

Taking students' ideas seriously (TSIS). TSIS involves valuing and building upon students' intuitive understanding of mathematical concepts (Carpenter & Lehrer, 1999; Hiebert, 1997; Romberg & Kaput, 1999). For example, when students solve an unfamiliar yet meaningful math problem, they draw on their prior knowledge and experience. Their solution strategies and notations may seem inefficient or informal to an observer, but by eliciting and valuing students' initial solution strategies, teachers can connect student thinking to more efficient and abstract methods (Freudenthal, 1973, 1991; Gravemeijer & van Galen, 2003; Treffers, 1987).

Pressing students conceptually (PSC). PSC focuses on building connections between mathematical strategies and models and progressively formalizing those ideas and methods for solving problems (Carpenter & Lehrer, 1999; Forman, 2003; National Research Council, 2001; Siegler, 2003). For example, once students have had the chance

to work on their own solution methods, teachers press them to connect and compare methods, generalize them to new situations, and relate them to formal mathematical terms and conventions. It is through this process of connection and generalization that students move from their own informal methods to more formal and efficient strategies (Carpenter & Lehrer, 1999; Gravemeijer & van Galen, 2003).

Encouraging multiple strategies and models (EMMS). EMMS involves developing students' understanding of various models and approaches to solving problems (Dolk & Fosnot, 2006; National Council of Teachers of Mathematics, 2000; Romberg & Kaput, 1999). When students generate, evaluate, and utilize different mathematical strategies and models, they recognize there are many ways to solve problems and represent solutions (Bruner, 1964). In addition, different strategies and models highlight different aspects of the mathematics, and thus examining the same problem through different lenses deepens students' overall understanding of the topic.

Addressing misconceptions (AM). AM involves using students' mistakes and misconceptions as valuable tools to build mathematical understanding (Borasi, 1987, 1994; Bray, 2013; Gooding & Stacey, 1992). Making mistakes and learning from them is an integral part of doing mathematics at any level. But mistakes often recur even after teachers demonstrate a correct procedure because they stem from deeper mathematical misconceptions. By being aware of why and how misconceptions develop and taking the time to address misconceptions through models and discussion, teachers can move students to a deeper level of understanding that precludes such mistakes. Additionally, mistakes can be opportunities for students to engage in justification, evaluation, and inquiry (Borasi, 1987).

Focusing on the structure of the mathematics (FSM). FSM involves facilitating students' understanding of fundamental, or structural, mathematical concepts (e.g., decomposing and composing, units and unitizing, equivalence). Many teachers and their 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 one is looking at 2nd or 11th grade. Focusing on these structures allows students to build understanding of and establish connections between these fundamental concepts and the particular topic being studied (National Council of Teachers of Mathematics, 2000; National Governors Association & Chief Council of State School Officers, 2011). 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.

Social-Constructivist and Transmission-Based Instructional Practices

In addition to examining the breadth of the mathematics instructional practice construct, it is important that surveys address the various perspectives within that domain. Teachers' instructional practice is often polarized in the mathematics instructional practice survey literature as either teacher-centered versus student-centered (e.g., Swan, 2006) or given other similar labels. While these labels simplify interpretation of the constructs under examination, they place the practices associated with each construct on opposite ends of a continuum without recognizing that teachers can effectively use both instructional approaches, depending on the situation or topic. For example, in a data unit, a teacher might utilize an exploratory activity and classroom discussion to generate ideas about how to represent and analyze the heights of students in the class. Later in the unit, the same teacher might directly instruct students about how to create a box and whisker plot, in line with mathematical conventions and norms. This direct instruction might then be followed by small-group discussions comparing different formal representations. Thus, rather than being at odds, transmission-based and social-constructivist learning may in fact be used simultaneously, and it would in fact be worrying if only one or the other were in constant practice. In addition, some labels themselves, for example "student-centered," tend to legitimize particular instructional practices, while others, such as "teacher-centered," are deprecating. If one of the purposes of measuring instructional practice is to determine the relationship of instructional practice to important variables of interest, it is necessary to use labels that accurately describe the trait under examination while also allowing for ease of interpretation of the meaning.

Based on our professional development, we wanted a survey that would allow teachers to realistically describe actual classroom practice without obviously "penalizing" a certain type of practice. We borrowed from Swan's (2006) framing of instructional practice as arising from different beliefs about how students learn. We utilized transmission-based (Cobb, 1988; Stipek, Givvin, Salmon, & MacGyvers, 2001) and social-constructivist (Cobb & Yackel, 1996) learning theory constructs to identify instructional practices and terminology that could arise in or describe a classroom based on each of these learning theories. We used those terms to frame instruction typically associated with teacher- and student-centered instructional practices, respectively. We recognize that these terms are typically associated with learning theories

and can manifest in multiple ways with regard to instructional practice. The DMT components provide a specific framework for instruction built upon these theories.

The next section describes the survey development process, addressing evidence of content validity. This is followed by the methods and results for Study 1, which examines the evidence of internal structure and capacity for the survey instrument to evaluate change in pedagogy by analyzing data from participants who took a 3-credit professional development course. Last, the methods and results for Study 2 examine the relationship of the survey constructs to observations of instructional practice utilizing data from participants involved in a 3-year professional development grant.

Survey Item Development

Our survey framework utilized three perspectives of instructional practice to ensure we measured the depth and breadth of the domain. The first perspective involved the use of the five domains within the DMT framework previously described (see the column headings in Table 2). This was followed by examining each DMT domain through the perspective of student practices, teacher practices, and classroom tasks and activities to address the depth within each domain (see the row headings in Table 2). Last, the transmission-based (T) and social-constructivist (S-C) learning theory perspectives were operationalized through survey items written to address each perspective. Table 2 provides an overview of the theoretical framework used in the initial development of survey items. We anticipated that the constructs of social-constructivist and transmission-based practices would emerge from the data.

Initially, one to three items were placed in each of the 30 cells of the framework. A total of 74 items were constructed. For example, in the domain of taking students' ideas seriously, a social-constructivist item with a student perspective was, "Students are encouraged to discuss their mathematical ideas in pairs, small-group, and/or whole-class discussions." A transmission item with a teacher perspective was, "I demonstrate for the class the correct way to use a particular procedure or model before they start solving problems."

To help establish the content validity of the instructional practice scale, a panel of six university-level mathematics education professors and professional developers analyzed the items. Any items deemed to inaccurately measure the construct domain were revised or removed. Once the initial development and review process was completed, the process took on a cyclical nature: survey item administration, analysis of the data in relation to variables of interest and psychometric properties, and finally revision and review. This process occurred three times and led to the refinement of the initial set of 74 items to 30 items, one addressing each cell within the framework.

Study 1

Study 1 focused on providing validity evidence related to internal structure and relationship to other variables. The research questions guiding Study 1 were:

1. Which items most strongly correlate to the constructs of social-constructivist and transmission-based instructional practice?

Table 2
Theoretical Framework to Address the Depth and Breadth of Classroom Mathematics Instructional Practice. T = Transmission, S-C = social-constructivism.

	Developing Mathematics Thinking Framework				
Perspective	Take students' ideas seriously	Encourage multiple models and strategies	Press students conceptually	Address misconceptions	Focus on the structure of mathematics
Student	T item(s)	T item(s)	T item(s)	T item(s)	T item(s)
	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)
Teacher	T item(s)	T item(s)	T item(s)	T item(s)	T item(s)
	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)
Tasks & activities	T item(s)	T item(s)	T item(s)	T item(s)	T item(s)
	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)	S-C item(s)

2. Using a separate sample, do the survey items correspond to their intended constructs?
3. What is the relationship between the constructs of social-constructivist and transmission-based instructional practice?
4. Do the survey scale scores for the two constructs have the capacity to evaluate change in intended practice as a result of professional development?

Data Source

The data for the present study came from participants enrolled in a three-credit, 45-hour mandated mathematics professional development course for kindergarten–12th grade teachers and administrators called Mathematical Thinking for Instruction (MTI). The course focuses on enhancing teachers' knowledge of mathematics, their understanding of how students best learn mathematics, and their ability to teach mathematics effectively. MTI instructors utilize the DMT framework—(1) taking students' ideas seriously, (2) encouraging multiple strategies and models, (3) pressing students conceptually, (4) focusing on the structure of the mathematics, and (5) addressing misconceptions, both implicitly and explicitly throughout the course. The MTI course uses number and algebra topics as the basis for understanding and applying the DMT framework. The professional development activities utilize participants' thinking, strategies, and models as the basis for discussion, thus modeling "taking students' ideas seriously" through both intensive tasks such as examining dividing fractions (e.g., Hughes, Brendefur, & Carney, 2015) or smaller investigations such as exploring multiplication facts through arrays. The ensuing group discussion is modeled around the DMT framework by pressing connections between participants' models and strategies and explicating the structure of the mathematics and potential misconceptions. This is often followed by explicit discussion of applications of the content and DMT framework to the K–12 classroom.

Survey Administration

Participants received a link to the instructional practice survey via email following course completion during the spring and summer of 2013. While the majority of course participants were teachers actively instructing in mathematics, the group also included (1) administrators, such as principals, district office personnel, and superintendents, (2) special education teachers who may or may not teach mathematics on a regular basis, and (3) all K–8 certified personnel, which included many middle school teachers of nonmathematics content. Individuals who did not teach mathematics on a regular basis were eliminated from the analysis, as indicated in the last column of Table 3.

Participants were asked to respond to the survey items following the completion of the MTI course (referred to as *after*) and through retrospective analysis of their practice prior to participation in the course (referred to as *prior*). Retrospective analysis was utilized because participants were likely to have a different understanding of the survey item wording and constructs following participation in the MTI course (Aiken & West, 1990; Lam & Bengo, 2003). Our approach of reporting postcourse and retrospective precourse instructional practice is the more conservative approach when compared with other forms of retrospective analysis that specifically measure change in practice (Lam & Bengo, 2003).

We had an overall survey response rate of 85.8%, with a total of 798 applicable course participants who taught mathematics on a regular basis completing the course and end-of-course survey during the spring and summer of 2013. For each participant, the data set contained 30 survey items with responses related to their instruction *prior* to the MTI course and the same 30 items with responses related to their instruction *after* the MTI course. The survey response scale ranged from 1 = Never to 7 = Daily (1= Never, 2 = 2–3 times per year, 3 = Once per month, 4 = 2–3 times per month, 5 = Once a week, 6 = 2–3 times per week, 7 = Daily). Given that the MTI course could span from 1 to 10 weeks and our survey scale spanned frequency of practice from daily to 2–3 times per year, what was likely captured in participants' responses *after* the MTI course was their intended, instead of actual, frequency of practice.

The survey items were grouped according to three perspectives: student, teacher, or task/activities. This design was intentional because we did not expect these aspects of the framework to emerge as constructs (i.e., latent variables) in the initial analysis. Within a survey page, the

Table 3
MTI Course and Survey Participants from Spring and Summer 2013 MTI Courses

Course level	Course participants	Survey participants	Applicable participants
K–3	933	828	497
4–8	550	429	229
6–12	141	136	72
Total	1624	1393	798

item stem specific to that perspective was placed at the top of the page with all related survey items below. The item order was randomized for each survey participant. The purpose of mixing construct items both within and across the pages of the survey was to ensure the constructs still held together when intentionally mixed. An example from the tasks/activities perspective page of the survey is provided in **Appendix A**.

Survey Refinement

As described previously, our survey development process resulted in 30 items: 15 designed to address social-constructivist practices and 15 designed to address transmission-based practices. The reality of survey development is that despite our best efforts, these items may or may not be good indicators of these two constructs. Part of the survey development process involves conducting statistical tests (e.g., exploratory factor analysis and confirmatory factor analysis) to determine which items are most highly correlated with (i.e., best measure) the constructs under examination. Once it has been verified that items cleanly measure a construct, it is more appropriate to find a total or average of those items as a representation (or operationalization) of that construct, often referred to as a scale score. Keep in mind that this numeric score still has a certain amount of measurement noise or variability (i.e., it does not perfectly measure the construct), but is a clearer representation of a unidimensional construct than if all the original items had been retained to create the scale score.

The purpose of the following analyses is to examine the internal structure of a survey of mathematics instructional practice. Based on our a priori theoretical framework, we anticipated latent variables around the constructs of socialist-constructivist and transmission-based learning theories. Confirmatory factor analysis (CFA) was utilized to examine the validity of the hypothesized factor structure. However, exploratory factor analysis (EFA) was initially utilized as a starting point to identify mis-fit items or potential subconstructs within the anticipated latent variables. The MTI course data set (n = 798) was randomly divided into two datasets so the two stages of analysis could be performed on independent samples. Following examination of the factor structure, the entire data set was analyzed to determine whether changes could be detected in the latent variables from the *prior* vs. *after* perspectives.

Results

Exploratory factor analysis (EFA) and item analysis. Exploratory factor analysis was used to determine which survey items did the best job of measuring (or were most highly correlated with) our constructs of social-constructivist and transmission-based practice. The technical write-up, factor loadings, and summary statistics are presented in **Appendix B**. Eight of the 15 items were identified as cleanly measuring the construct of social-constructivist practice, and 6 of the other 15 items were identified as cleanly measuring the construct of transmission-based instructional practice. The items and associated constructs are presented in Table 4, and an example of these items in a survey format is provided in **Appendix C**.

In addition to the exploratory factor analysis of item correlations, means and standard deviations were computed. These are provided in **Appendix D** for *prior* (Table D1) and *after* (Table D2) MTI course participation. The items within each identified construct significantly correlated to one another. In other words, social-constructivist items correlated significantly with one another, and transmission items correlated significantly with one another. In addition, internal consistency was examined for each scale and was found to be good to excellent for all scales; social-constructivist (*prior* α = .91; *after* α = .90) and transmission-based (*prior* α = .84; *after* α = .86).

Confirmatory factor analysis (CFA). The purpose of confirmatory factor analysis is to determine whether the sample data do a reasonable job of matching a hypothesized model. The major components of our hypothesized model included (a) two latent variables or constructs associated with social-constructivist (soccon) and transmission-based (trans) practice, and that (b) each survey item would correlate to the latent variable it was designed to measure and not correlate to the other latent variable. Our analyses indicated that the sample data do a reasonable job of matching a hypothesized model, confirming the results of our EFA (detailed CFA findings in **Appendix E**). Second, the construct of social-constructivist instructional practice was for the most part uncorrelated to transmission-based instructional practice. The lack of a moderate to strong negative correlation between the two constructs indicates our survey participants view these as distinct and unrelated constructs within mathematics instructional practice, rather than as two ends of a single instructional spectrum.

Capacity to evaluate change. Following the EFA and CFA analyses, the entire data set was combined and evaluated for the capacity of the scale scores for social-constructivist and transmission-based practice to capture change as a result of professional development. While a lack of change from *prior* to *after* would not invalidate the survey's capacity to capture change—the lack of change could be due to the professional development itself—the

Table 4
Survey Items and Associated Constructs Retained for CFA

Construct	Survey items (DMT framework element)	Item label for CFA
Social-constructivist items	Facilitate small-group or whole-class discussion on student thinking (TSIS_T)	SC1
	Are based on their potential to encourage discussions of students' mathematical ideas (TSIS_A)	SC2
	Emphasize the use of multiple models for recording and communicating student thinking (EMRS_T)	SC3
	Solve problems that allow for several different approaches (EMRS_S)	SC4
	Facilitate discussion about underlying mathematical concepts (e.g., composing or decomposing number; FSM_T)	SC5
	Are selected because they provide opportunities for students to explain the mathematics behind an answer (FSM_A)	SC6
	Encourage discussion of the connections between various models and strategies (PSC_T)	SC7
	Analyze the connections between various models and procedures (PSC_S)	SC8
Transmission items	Demonstrate for the class the correct way to use a particular procedure or model before they start solving problems (TSIS_T)	T1
	Present one standard method of solving a task or performing an algorithm (EMRS_T)	T2
	Explain the steps to a procedure or algorithm when I introduce new topics (FSM_T)	T3
	Take notes on how to perform each step in a procedure or algorithm (FSM_S)	T4
	Learn by copying down examples from a teacher demonstration (PSC_S)	T5
	Avoid student errors and misconceptions when a topic is first introduced by explaining how to solve a problem before they start (AM_T)	T6

demonstrated ability to capture change from a professional development course that has already successfully demonstrated significant changes in teachers' knowledge, beliefs, and self-efficacy (Carney, Brendefur, Thiede, et al., 2014) assists in the initial validation efforts. Scale scores were constructed by finding the mean of the eight items for social-constructivist and six items for transmission-based practices, eliminating those individuals missing more than one item score within a scale. Appendix F provides histograms for each of the variables from the perspective of *prior* and *after* MTI course participation. The histogram for social-constructivist practice *after* MTI course participation indicates a potential issue with the frequency of practice scale; there is a ceiling effect on reporting particular types of practice.

A Wilcoxon signed-rank test showed that participants' self-reported, retrospective analysis of their instructional practice indicated significant changes across both constructs; social-constructivist (Z = 22.718, p < .001) and transmission (Z = 20.072, p <.001). The scores can be interpreted within the original scale metric; 1 = Never, 2 = 2–3 times per year, 3 = Once per month, 4 = 2–3 times per month, 5 = Once a week, 6 = 2–3 times per week, 7 = Daily. The median for social-constructivist practice was 4.6 (n = 847) for *prior* and 6.6 (n = 784) for *after* MTI course participation. On the original metric, this would indicate that prior to the MTI course, teachers reported engaging in social-constructivist practices less than once a week; following the MTI course, they indicated a shift in social-constructivist practice to more than 2–3 times per week. The median for transmission-based instructional practice demonstrated similar changes to the social-constructivist variable but in the opposite direction, with 5.7 (n = 857) *prior* to and 3.6 (n = 772) *after* the MTI course. On the original metric, this indicates a shift from conducting transmission-based practice more than once a week before the course to less than 2–3 times per month *after* the course.

Study 2

Study 2 focused on providing validity evidence related to the relationship between variables of interest: What is the relationship between teachers' self-reported instructional practice survey scores (both social-constructivist

and transmission-based practice scale scores) and scores based on observations of their practice?

Data Source

The data for study two consisted of a sample of 39 fourth- through eighth-grade teachers, who were in their 3rd year of a professional development grant around the DMT professional development framework (Brendefur, Thiede, Strother, Bunning, & Peck, 2013). The MTI course, previously described, was an outgrowth of this much more intensive and sustained professional development project, which involved summer coursework, planning meetings, and in-class support over 3 years. As part of the professional development, teachers were observed once during the fall of the 3rd year of the project, and they completed the instructional practice survey in the winter of the same year. Teachers' classroom practice was evaluated using a DMT observation instrument built around the DMT framework (Carney, Brendefur, & Hughes, 2014). Similar to our findings with existing instructional practice surveys, existing observation measures (e.g., Mathematical Quality of Instruction, Reformed Teaching Observation Protocol, and Instructional Quality Assessment) framed classroom instructional practice in similar ways but were different enough that they did not capture the full breadth of change we envisioned as a result of our professional development work.

The five domains of the DMT framework provided the structure for the observation instrument development. Each DMT domain was measured with four items. Each item had an overall descriptor followed by specific item descriptors for each score, which ranged from 1–5. The four item scores across each of the five domains were compiled into an overall average for instructional practice on a scale of 1–5, representing the teachers' level of engagement in the elements of the DMT framework (1 = unaffected, 2 = developing, 3 = engaged, 4 = accomplished, 5 = reflective). The small number of items per domain did not support individual analyses of each domain. The DMT instrument has been shown to have high internal reliability (α = .89 to .98).

Data Analysis

Pearson's *r* was calculated to examine the relationship between the constructs in the instructional practice survey and the teachers' scores from the DMT classroom observation instrument. High scores on the observation instrument indicate teachers engaging in the DMT instructional practices at higher levels of quality and quantity, whereas high scores on the survey instrument indicate teachers' reporting high frequency of engagement in the DMT instrument practices.

Results

The correlation analysis indicated moderate to high levels of correlation between the DMT observation instrument and the two constructs from the instructional practice survey, social-constructivist and transmission-based practice (see Table 5). The social-constructivist variable positively correlates ($r = .37$, $n = 39$, $p < .05$) and the transmission-based variable negatively correlates ($r = –.45$, $n = 39$, $p < .05$) with teachers' score on the DMT instructional practice observation instrument. This indicates a relationship between teachers' self-reports of frequency of instructional practice with observations of their practice conducted by others. In other words, this provides

Table 5
Pearson's r *and Descriptive Statistics for Instructional Practice Survey Factors 1–3 and Observation Score 4 (*n *= 39)*

Variables	1	2	3
1. Social-constructivist	1		
2. Transmission-based	-0.26	1	
3. DMT observation score	.37*	-.45**	1
M	5.97	3.74	56.92
SD	0.66	1.53	18.06
Range	1–7	1–7	20–100

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

initial evidence that the survey scale scores reflect teachers' relative frequencies of these instructional practices. Scatterplots of these data are provided in **Appendix G**.

Discussion and Implications

The initial psychometric evaluation of the survey items revealed several findings. First, the constructs of social-constructivist and transmission-based instructional practices were supported by both the exploratory and confirmatory factor analyses providing evidence of internal structure. In addition, the lack of a moderate to large negative correlation relationship between the social-constructivist and the transmission-based constructs is important to consider given the opposing ways these variables are often presented in the literature and treated in survey measures. The analysis of the retrospective change in instructional practice from *prior* to *after* course participation provides initial evidence that the survey scales capture change as a result of particular professional development activities but also reveal an issue with a ceiling effect in the frequency scale. Last, Study 2 provides initial concurrent validity evidence of the relationship between teachers' self-reported frequency and observations of instructional practice. The following section discusses each of these specific findings followed by more general implications and conclusions.

The two constructs of social-constructivist and transmission-based practice were supported by the exploratory and confirmatory factor analyses. The social-constructivist based practice scale was made up of eight items with two items each coming from four of the five domains from the DMT framework: four from a teacher perspective, two from a student perspective, and two from a tasks and activities perspective. Based on the extent to which the final survey items reflected the original framework, we have reasonable content validity evidence that the final scale reflects the depth and breadth of the social-constructivist based aspects of the DMT framework. In other words, our original theoretical framework from Table 2 was reasonably represented by the eight items remaining following the EFA. However, the lack of items representing the *addressing misconceptions* domain is concerning. This could be interpreted as a result of a lack of consistent focus within our professional development activities in clearly articulating these ideas. However, there may be other potential causes for the lack of cohesiveness with these items and the social-constructivist based practice scale as operationalized (e.g., participants' readiness to benefit). Future scale development work will focus on investigating the *addressing misconceptions* domain and its relationship to the other constructs.

The transmission-based scale consists of six items, one item from each of the DMT framework domains except for *focusing on the structure of the mathematics*, which has two items. The majority of the items came from a teacher-based perspective (four items), with the other two items coming from the student-based perspective. Again, this provides reasonable content validity evidence that the survey items appear to clearly capture a transmission-based perspective of mathematics instruction. In other words, the six items representing transmission-based practice reasonably represent the breadth of the theoretical framework. However, further scale development will focus on determining whether there are potential tasks or activities that relate specifically to this construct.

The lack of a moderate to strong negative correlation between the transmission and social-constructivist items is a particularly interesting finding. These constructs are often treated as opposing ideas within the mathematics education literature and in particular in measures of teachers' instructional practice. For example, Swan (2006) reverse scored the student-centered items and placed them on the same scale as the teacher-centered items. Ross and colleagues (2003) used a similar design. This assumes unidimensionality and a polar relationship, which we found not to be the case. In other words, it appears from our data that social-constructivist and transmission-based instructional practice are not opposite ends of a continuum and should be examined as two separate variables. Whether our results are particular to our sample or whether this is a consistent finding across teachers requires further investigation but should make researchers consider analyzing the structure of data from items that are typically reverse scored.

The evidence related to the ability of the items to measure change in instructional practice as a result of professional development shows promise. Our findings indicate a significant change in the participants' perceptions of their instructional practice from *prior* to *after* using a retrospective model. Further investigation needs to occur to determine how closely survey responses match observed changes in practice and whether the internal structure holds beyond the bounds of our particular professional development project. In addition, the ceiling effect in the frequency scale may indicate a need to expand the current scale in future studies.

Study 2 examined the concurrent validity between the self-reported survey constructs and observations of instructional practice. The moderately to strongly significant correlations in the expected directions provide promising validity evidence regarding the accuracy of teachers' self-reported frequency of instructional practice. In other

words, it provides evidence that teachers' self-reported instructional practices are moderately correlated with their actual practices. The DMT observation instrument is designed to capture the quality of mathematics instructional practice, while our survey instrument is designed to capture the frequency of those practices. The fact that we were able to capture moderate to strong relationships between these variables with related but different scale foci is promising. The evidence of this relationship should continue to be examined in future studies.

Lastly, what does all this mean to mathematics teacher educators looking for a survey scale to measure instructional practice? The validity evidence from each survey scale must be examined in relation to the context in which it will be used, such as the specific goals of the professional development program, how the constructs within a survey are defined and operationalized, and the complexities of creating and validating a survey scale. As a research community, mathematics teacher educators need a range of scales to select from and need to understand the various pros and cons of each. The following ideas are provided to stimulate thinking about various considerations when selecting a survey scale.

The *TIMSS-Engage and Horizon-Reform* survey scales are particularly useful for researchers who want to compare their results with an international or national sample. However, researchers using these scales need to closely examine how the scales conceptualize instructional practice to determine whether that conceptualization matches the research project's needs. The *Swan-Practices* and *Ross-Standards* scales provide a much larger item bank in their survey scales and therefore may provide a broader picture of instructional practice. However, researchers who use these scales may want to utilize factor analysis to examine the unidimensionality of these scales. These scales offer more detailed evidence of the relationship of their scale scores to other variables of potential interest. However, this information needs to be carefully considered in terms of its generalizability to other situations.

Our instructional practice instrument provides separate scales for social-constructivist and transmission-based variables. This allows researchers to evaluate the level of each variable separately within their sample. In addition, initial evidence supports the use of this instrument on a large scale to investigate change in instructional practice as a result of professional development focused on aspects of the DMT framework.

Through the lens of our own professional development project, our current results indicate we increased the frequency with which course participants engaged in social-constructivist practices and decreased the frequency with which participants engaged in transmission-based practices. The results of Study 2 provide initial support for the claim that these self-reported changes in practice translated into actual classroom practice. Given the large-scale nature of our project—to date over 12,000 participants have completed the course—this provides promising evidence that high-quality, large-scale mandated professional development has the potential to shift classroom practice.

Additional validation work needs to be conducted to determine the usefulness of the instrument across varied settings (e.g., outside of Idaho) and participants (e.g., preservice teachers) and the relationship between the survey scale scores to other variables and/or constructs of importance (e.g., student achievement). For example, examining the relationship between teachers' reported frequency of instructional practice for each construct in relation to measures of student achievement or socioemotional well-being has the potential to provide quantitative evidence for the frequency with which teachers should engage in different types of instructional practice. The current trend appears to support increasing social-constructivist practices and decreasing transmission-based practices, but perhaps how these two modes interplay in the classroom might give us more useful information. In addition, our focus on frequency versus quality of instructional practice should be kept in mind. How might this impact findings? We welcome careful use, modification, and further study of this instrument and hope it serves to spark further discussion around ideas of measuring practice on a large scale.

References

Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments self-report pretest biases. *Evaluation Review*, *14*(4), 374–390.

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*. Chapel Hill, NC: Horizon Research, Inc.

Boaler, J. (2005). Promoting equity in mathematics classrooms–important teaching practices and their impact on student learning. Paper presented at the 10th International Congress on Mathematics Education, Copenhagen, Denmark.

Borasi, R. (1987). Exploring mathematics through the analysis of errors. *For the Learning of Mathematics*, *7*(3), 2–8.

Borasi, R. (1994). Capitalizing on errors as "springboards for inquiry": A teaching experiment. *Journal for Research in Mathematics Education*, *25*(2), 166–208.

Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response-shift bias. *Educational and Psychological Measurement*, *44*(4), 781–804.

Bray, W. S. (2013). How to leverage the potential of mathematical errors. *Teaching Children Mathematics*, *19*(7), 424–431.

Brendefur, J. L., Carney, M. B., Hughes, G., & Strother, S. (forthcoming). 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), 62–75.

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

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

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

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2014). Statewide mathematics professional development: Teacher knowledge, self-efficacy, and beliefs. *Educational Policy*, 1–34, doi:10.1177/0895904814550075

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: cognitively guided instruction*. Portsmouth, N.H.: Heinemann.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. A. Romberg (Eds.), *Mathematics Classrooms that Promote Understanding* (pp. 19–32). Mahwah, N.J.: Lawrence Erlbaum Associates.

Cobb, P. (1988). The tension between theories of learning and instruction in mathematics education. *Educational Psychologist*, *23*(2), 87–103.

Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist*, *31*(3–4), 175–190.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*(2), 166.e7–166.e16. http://dx.doi.org/10.1016/j.amjmed.2005.10.036

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1–9.

D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievements. *School Effectiveness and School Improvement*, *11*(2), 197–235.

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, *26*(1), 1–22.

Dolk, M., & Fosnot, C. T. (2006). *Minilessons for operations with fractions, decimals, and percents addition and subtraction*. Portsmouth, NH: Heinemann.

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching: International Research and Pedagogy*, *15*(1), 13–33.

Forman, E. A. (2003). A sociocultural approach to mathematics reform: Speaking, inscribing, and doing mathematics within communities of practice. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, the Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, the Netherlands: Kluwer.

Gooding, A., & Stacey, K. C. (1992). *Communication and learning in small group discussion*. Paper presented at the Seventh International Congress of Mathematical Education, Quebec, Canada.

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: National Council of Teachers of Mathematics.

Hiebert, J. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hughes, G., Brendefur, J., & Carney, M. (2015). Reshaping teachers' mathematical perceptions: Analysis of a professional development task. *Mathematics Teacher Educator*, *3*(2), 116–129.

Institute of Education Sciences & National Science Foundation. (2013). *Common guidelines for education research and development*. Washington, DC: Institute of Education Sciences.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement*. Westport, CT: Rowman & Littlefield.

Lam, T. C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, *24*(1), 65–80.

Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). Methods and procedures in TIMMS and PIRLS 2011. Boston, MA: TIMMS & PIRLS International Study Center, Boston College.

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

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Governors Association & Chief Council of State School Officers. (2011). *Common Core State Standards for Mathematics*. http://www.corestandards.org

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Romberg, T. A., & Kaput, J. J. (1999). Mathematics worth teaching, mathematics worth understanding. In E. Fennema & T. A. Romberg (Eds.), *Mathematics classrooms that promote understanding*. Mahwah, N.J.: Lawrence Erlbaum Associates.

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.

Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. I. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*, 213–226.

Swan, M. (2006). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research in Education*, *75*, 58–70.

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

Wubbels, T. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, *8*(1), 47–58.

Authors

Michele B. Carney, PhD, 1910 University Drive, Boise State University, Boise, ID 83725-1745; michelecarney@boisestate.edu

Jonathan L. Brendefur, PhD, 1910 University Drive, Boise State University, Boise, ID 83725-1745; jonathanbrendefur@boisestate.edu

Gwyneth R. Hughes, 1910 University Drive, Boise State University, Boise, ID 83725-1745; gwynethhughes@boisestate.edu

Keith Thiede, 1910 University Drive, Boise State University, Boise, ID 83725-1745; keiththiede@boisestate.edu

Appendix A: Mathematical Instructional Practice

Mathematics Instructional Practice - Tasks and Activities

Indicate for each statement the frequency you engage in the particular instructional practice both **prior** to and **after** participation in the MTI course.

Classroom tasks and activities:

	BEFORE Participation in MTI	AFTER Participation in MTI
Are selected because they provide opportunities for students to explain the mathematics behind an answer.		
Are directed by the sequence of the textbook		
Are based on the connections that can be made between various models and algorithms		
Involve the intentional presentation of solution strategies containing misconceptions and/or mistakes for student to diagnose and correct		
Are selected because the problem's context (or situation) elicits a particular model or models		
Are selected because they allow students repeated practice to learn a procedure		
Focus on repeated practice of a model or procedure		
Focus on mathematical procedures in order to avoid confusion and prevent student errors		
Primarily focus on learning a particular procedure or algorithm		
Are based on their potential to encourage discussions of students' mathematical ideas		

A drop-down box for each question lists the following options:

- Never
- 2¬–3 times per year
- Once per month
- 2–3 times per month
- Once a week
- 2–3 times per week
- Daily

(**Return to page 101**)

Appendix B: Exploratory Factor Analysis

Exploratory factor analysis. For each anticipated latent variable, an EFA using principal components extraction method with varimax rotation was conducted related to instruction both *prior* to and *after* the MTI course (using SPSSv21). The first EFA analyzed (1) 15 social-constructivist items measuring *prior* practice and (2) the same 15 social-constructivist items measuring *after* practice, and the second EFA analyzed (3) 15 transmission-based items measuring practice *prior* and (4) the same 15 items measuring transmission-based practice *after*.

The factor loadings and summary statistics are presented for the social-constructivist (Table B1) and transmission (Table B2) items. For social-constructivist, two factors were extracted with eigenvalues over one, 6.595 and 1.211 for *prior,* and three factors with eigenvalues over one, 5.895, 1.503, and 1.129 for *after*. Eight items had factor loadings above .40, loaded on the same component both *prior* and *after*, and had no cross-loading. These eight items were identified as measuring the latent variable social-constructivist practice and retained for analysis in the CFA.

For transmission, two factors were extracted with eigenvalues over one, 6.808 and 1.459 for *prior*, and three factors with eigenvalues over one, 6.946, 1.271, and 1.003 for *after*. Component one had six items with factor loadings over .40 and loaded only on component one both *prior* and *after* with no cross-loading. These six items were identified as measuring the latent variable of transmission-based instructional practice and retained for analysis in the CFA. Due to the dramatic drop in eigenvalues after component one, the remaining items were dropped from further analysis (Costello & Osborne, 2005).

Table B1
Rotated Component Matrix for Social-Constructivist Items

DMT framework element	***Prior* components**		***After* components**		
	1	**2**	**1**	**2**	**3**
TSIS_T	**.712**	-.063	**.734**	.033	.252
TSIS_A	**.736**	.190	**.750**	.068	.172
EMRS_T	**.768**	.026	**.731**	.066	.027
EMRS_S	**.797**	.072	**.705**	.158	.225
FSM_T	**.750**	-.055	**.750**	.159	.162
FSM_A	**.748**	.239	**.780**	.081	.019
PSC_T	**.748**	.067	**.631**	.117	.217
PSC_S	**.801**	.190	**.706**	.257	.168
EMRS_A	.159	.757	.161	.632	–.265
FSM_S	.576	.174	.212	.622	.249
PSC_A	.575	.381	.342	.682	.120
AM_T	.520	.002	.212	-.024	.783
AM_A	.573	.266	.252	.201	.735
AM_S	.670	.122	.574	.168	.441
Eigenvalues	6.595	1.211	5.895	1.503	1.129
Variance explained	44.0%	8.1%	39.3%	10.0%	7.5%

Table B2
Rotated Component Matrix for Transmission Items

DMT framework element	*Prior* components		*After* components		
	1	2	1	2	3
TSIS_T	**.679**	.378	**.786**	.231	-.076
EMRS_T	**.731**	.249	**.667**	.199	.087
FSM_T	**.717**	.355	**.798**	.231	-.015
FSM_S	**.528**	.120	**.644**	.299	.030
PSC_S	**.701**	.203	**.661**	.351	-.151
AM_T	**.716**	.296	**.755**	.183	.004
EMRS_S	.213	.795	.232	.777	.087
EMRS_A	.276	.743	.201	.832	.033
PSC_A	.222	.763	.232	.819	-.058
TSIS_A	.441	.269	.323	.395	.136
PSC_T	.543	.422	.564	.308	-.051
FSM_A	.692	.431	.560	.556	-.061
AM_A	.609	.483	.506	.601	-.166
AM_S	.331	.696	.505	.608	-.021
TSIS_S	-.553	.390	-.045	.032	.974
Eigenvalues	6.808	1.459	6.946	1.271	1.003
Variance explained	45.4%	9.7%	46.3%	8.5%	6.7%

TSIS = Taking students' ideas seriously; PSC = Pressing students conceptually; FSM = Focusing on the structure of mathematics; AM = Addressing misconceptions; EMRS = Encouraging multiple representations and strategies
A = Activities (and tasks); T = Teacher; S = Student

(**Return to page 101**)

Appendix C: Mathematics Instructional Practice Survey

Mathematics Instructional Practice Survey

Indicate for each statement the frequency you engage in the particular instructional practice.

As the classroom teacher, I:

	Daily	2-3 times per week	Once per week	2-3 times per month	Once per month	2-3 times per year	Never	Not applicable
Emphasize the use of multiple models for recording and communicating student thinking	○	○	○	○	○	○	○	○
Encourage discussion of the connections between various models and strategies	○	○	○	○	○	○	○	○
Facilitate discussion about underlying mathematical concepts (e.g., composing or decomposing number)	○	○	○	○	○	○	○	○
Present one standard method of solving a task or performing an algorithm	○	○	○	○	○	○	○	○
Facilitate small group or whole class discussion on student thinking	○	○	○	○	○	○	○	○
Explain the steps to a procedure or algorithm when I introduce new topics	○	○	○	○	○	○	○	○
Demonstrate for the class the correct way to use a particular procedure or model before they start solving problems	○	○	○	○	○	○	○	○
Avoid student errors and misconceptions when a topic is first introduced by explaining how to solve a problem before they start	○	○	○	○	○	○	○	○

Students:

	Daily	2-3 times per week	Once per week	2-3 times per month	Once per month	2-3 times per year	Never	Not applicable
Analyze the connections between between various models and procedures	○	○	○	○	○	○	○	○
Take notes on how to perform each step in a procedure or algorithm	○	○	○	○	○	○	○	○
Learn by copying down examples from a teacher demonstration	○	○	○	○	○	○	○	○
Solve problems that allow for several different approaches	○	○	○	○	○	○	○	○

Classroom tasks and activities:

	Daily	2-3 times per week	Once per week	2-3 times per month	Once per month	2-3 times per year	Never	Not applicable
Are based on their potential to encourage discussions of students' mathematical ideas	○	○	○	○	○	○	○	○
Are selected because they provide opportunities for students to explain the mathematics behind an answer.	○	○	○	○	○	○	○	○

(Return to page 101)

Appendix D: Descriptive Item Statistics

Table D1
Item Correlations, Means, and Standard Deviations for Prior to MTI Course Participation

Correlations *PRIOR*	Social-constructivist items								Behaviorist					
	SC1	**SC2**	**SC3**	**SC4**	**SC5**	**SC6**	**SC7**	**SC8**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
SC1	1													
SC2	.524**	1												
SC3	.512**	.535**	1											
SC4	.503**	.568**	.601**	1										
SC5	.501**	.446**	.600**	.511**	1									
SC6	.482**	.677**	.531**	.602**	.510**	1								
SC7	.515**	.491**	.626**	.556**	.519**	.501**	1							
SC8	.492**	.554**	.556**	.636**	.533**	.586**	.572**	1						
T1	-.146**	-.171**	-.189**	-.124*	-.135**	-.156**	-.140**	-0.07	1					
T2	-.252**	-.297**	-.389**	-.315**	-.255**	-.249**	-.297**	-.226**	.503**	1				
T3	-.142**	-.182**	-.177**	-.129*	-.120*	-.115*	-.175**	-0.09	.633**	.552**	1			
T4	-0.03	-0.08	-0.01	0.02	0.03	-0.02	-0.02	0.09	.327**	.301**	.369**	1		
T5	-.243**	-.272**	-.268**	-.259**	-.239**	-.207**	-.225**	-.127*	.487**	.473**	.526**	.438**	1	
T6	-.229**	-.220**	-.242**	-.263**	-.210**	-.263**	-.227**	-.225**	.578**	.511**	.545**	.314**	.515**	1
Mean	4.75	4.25	4.20	4.51	3.97	4.24	4.37	4.09	5.86	5.09	5.69	4.62	5.31	5.27
SD	1.90	2.08	2.06	1.95	2.16	2.04	2.06	2.05	1.60	2.08	1.72	2.25	2.04	2.06

Table D2
Item Correlations, Means, and Standard Deviations for After MTI Course Participation

Correlations *AFTER*	Social-constructivist items								Behaviorist					
	SC1	**SC2**	**SC3**	**SC4**	**SC5**	**SC6**	**SC7**	**SC8**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
SC1	1													
SC2	.508**	1												
SC3	.493**	.517**	1											
SC4	.507**	.594**	.562**	1										
SC5	.551**	.569**	.510**	.565**	1									
SC6	.530**	.689**	.520**	.541**	.550**	1								
SC7	.511**	.472**	.541**	.502**	.458**	.466**	1							
SC8	.487**	.589**	.522**	.619**	.580**	.576**	.476**	1						
T1	-.118*	-138**	-.157**	-0.05	-.144**	-.137**	-0.06	-0.07	1					
T2	-.102*	-0.09	-.188**	-0.04	-0.08	-0.07	-0.06	-0.02	.428**	1				
T3	-0.09	-0.09	-.158**	0.01	-0.09	-0.08	-0.09	0.00	.671**	.453**	1			
T4	-0.04	-0.03	-0.10	-0.03	-0.10	-0.05	-0.05	.107*	.459**	.398**	.502**	1		
T5	-.159	-.195	-.237	-.104	-.202	-.221	-.155	-0.06	.536	.473	.494	.607	1	
T6	-.134**	-0.08	-0.09	-0.03	-.157**	-.110*	-0.05	-0.07	.598**	.475**	.541**	.421**	.457**	1
Mean	6.41	6.28	6.36	6.36	6.30	6.29	6.39	6.21	3.96	3.03	3.99	3.78	3.64	3.09
SD	0.90	1.18	0.95	1.05	1.09	1.10	1.06	1.12	2.24	2.17	2.20	2.20	2.11	2.25

(Return to page 101)

Appendix E: Confirmatory Factor Analysis

Confirmatory factor analysis. The purpose of confirmatory factor analysis is to determine the goodness of fit between a hypothesized model and the sample data. Our hypothesized a priori structure is comprised of (a) two factors, social-constructivist (soccon) and transmission (trans), (b) with each item having a nonzero loading on the factor it is designed to measure and a zero loading on factors it is not designed to measure, and (c) the residuals for each item are uncorrelated with each other. The next stage of analysis involved using the second half of the randomly split data to examine the fit of this model (n = 399). The MLM estimator in Mplus7 was utilized due to issue with multivariate non-normality, requiring listwise deletion of missing data and reducing the sample size (n = 326 *prior*, n = 319 *after*). Our initial analysis indicated a potential model misspecification, and our model was adjusted post hoc to allow for error covariance between items T2 and T6, T4 and T5, and SC2 and SC6. Due to the conceptual similarity in the items, this adjustment still fit our initial framework. Figure E1 contains the full diagram results of the final CFA, and Table 5 provides the fit statistics. While there is no clear consensus regarding the indices that are most appropriate for model fit, our indices fit all the major recommendations (Byrne, 2013).

Table 5

Fit Statistics for Final CFA Model with MLM Estimator for Prior and After MTI Course Participation

Model	N	X^2	df	RMSEA	SRMR	CFI
Prior	326	131.30*	73	.049	.067	.970
After	319	101.56*	73	.035	.044	.972

* $p < .01$

Two key findings emerged from the CFA. First, the factor structure resulting from the EFA was supported by the CFA model. In other words, the survey item responses loaded on the latent factors as expected. Second, the construct of social-constructivist instructional practice was found to be only negligibly correlated to transmission-based instructional practice for *prior* r = -.18 and not significantly correlated for *after* r = -.16. The lack of a moderate to strong negative correlation between the two latent variables indicates our survey participants view these as distinct and unrelated constructs within mathematics instructional practice.

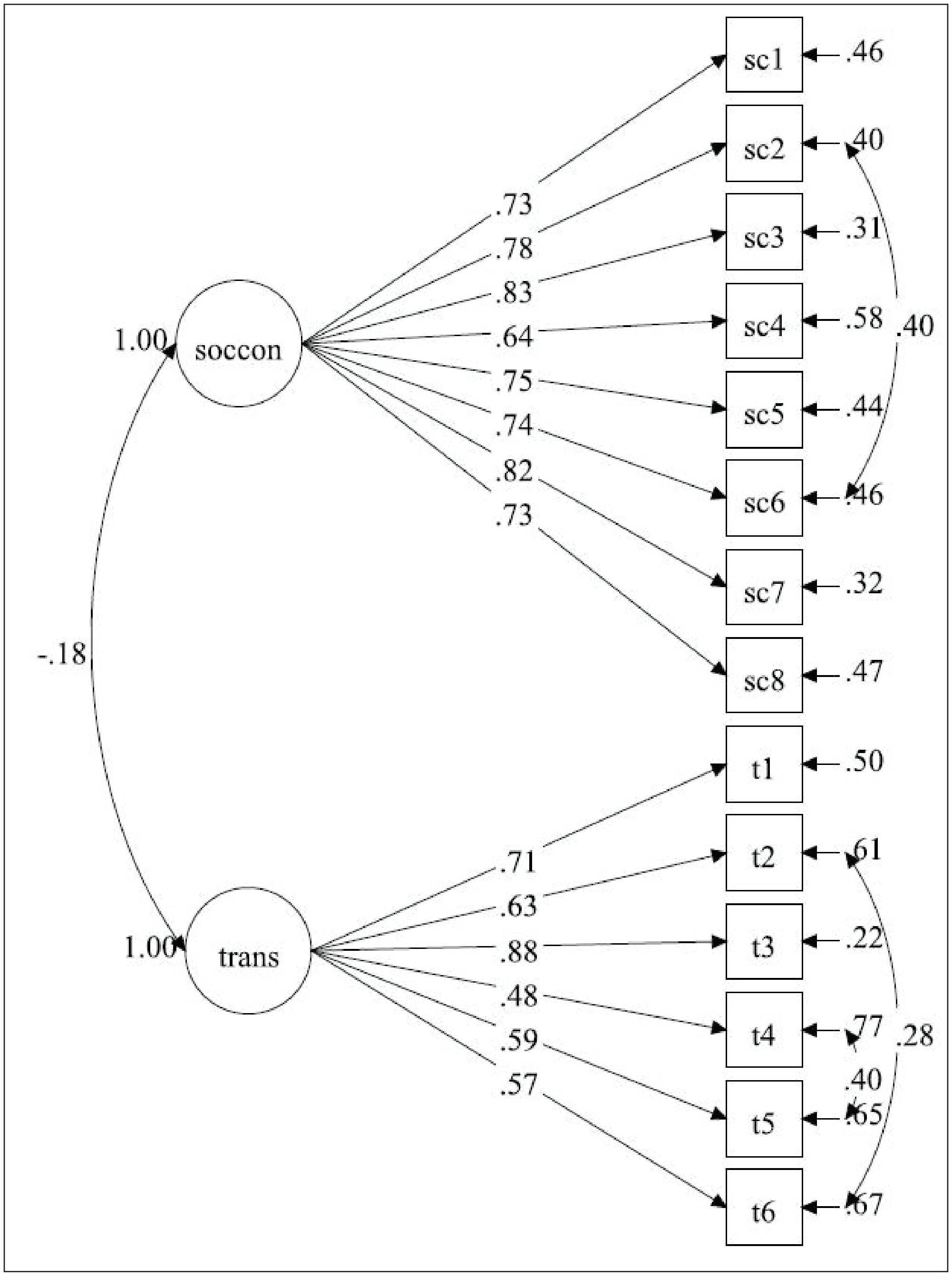

Figure E1. **CFA for responses related to instruction *prior* to the MTI course.**

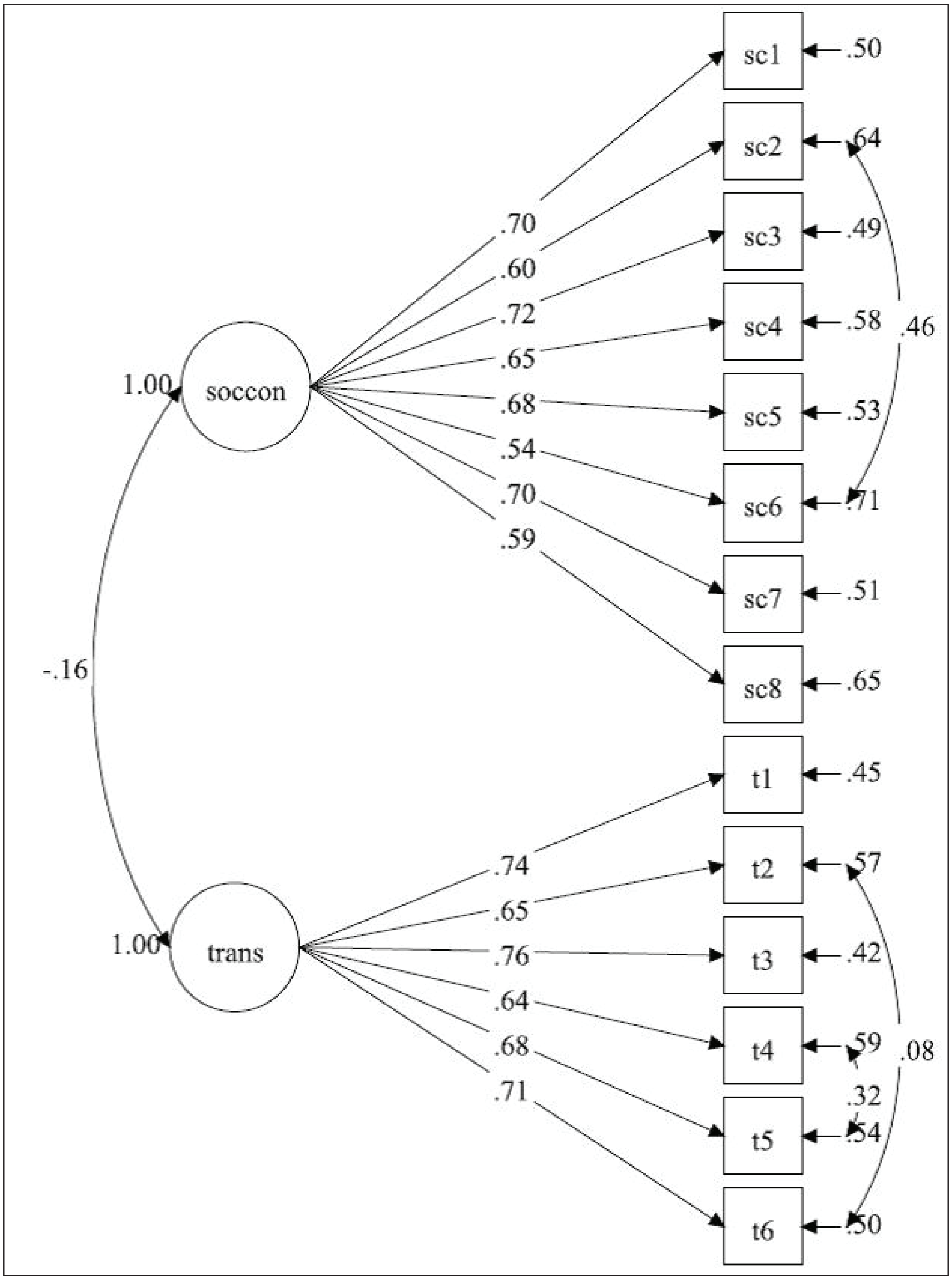

Figure E2. **CFA for responses related to instruction *after* the MTI course.**

(Return to page 101)

Appendix F: Histograms of Scale Scores

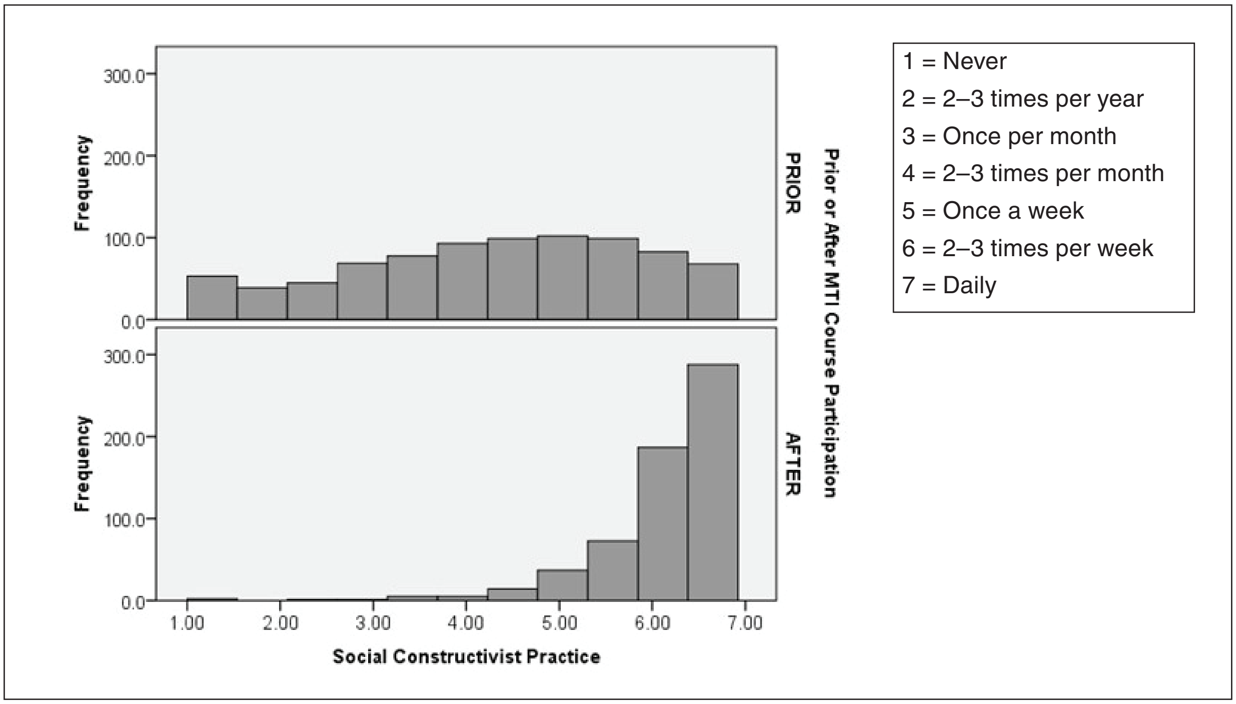

Figure F1. **Histograms of social-constructivist scale scores for *prior* and *after* MTI course participation.**

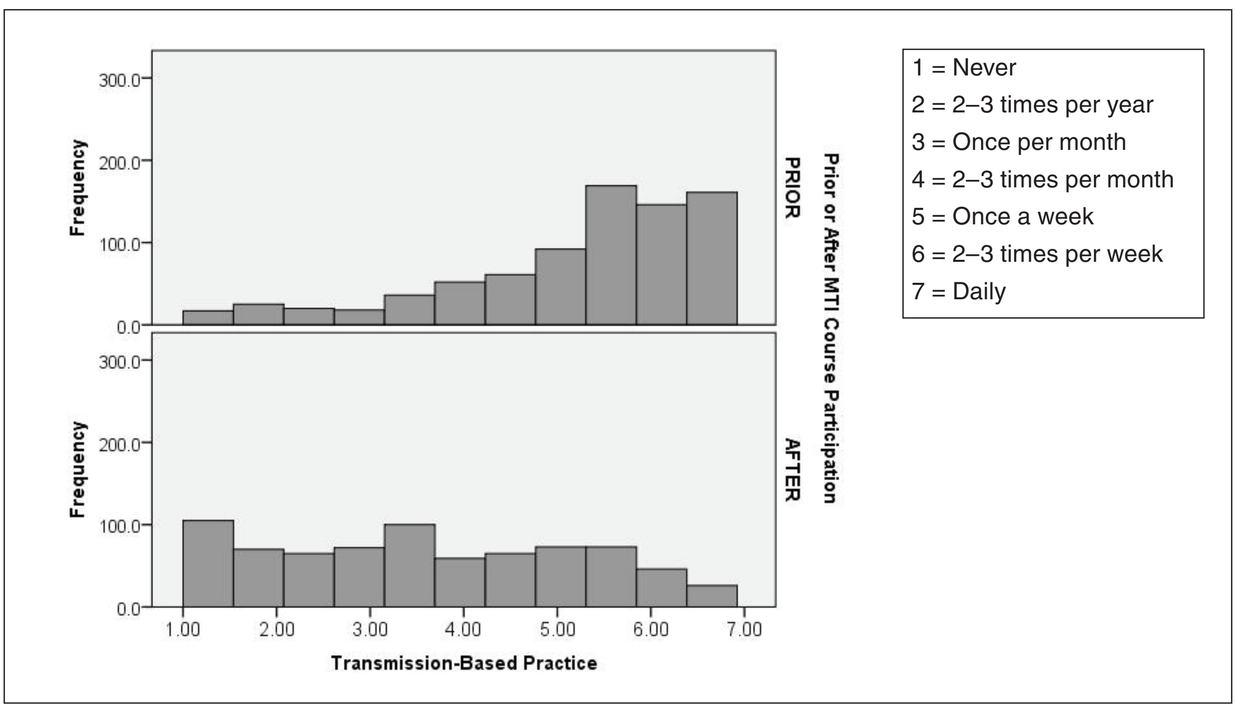

Figure F2. **Histograms of transmission-based scale scores for *prior* and *after* MTI course participation.**

(Return to page 102)

Appendix G: Correlation of Scale Scores to DMT Observation

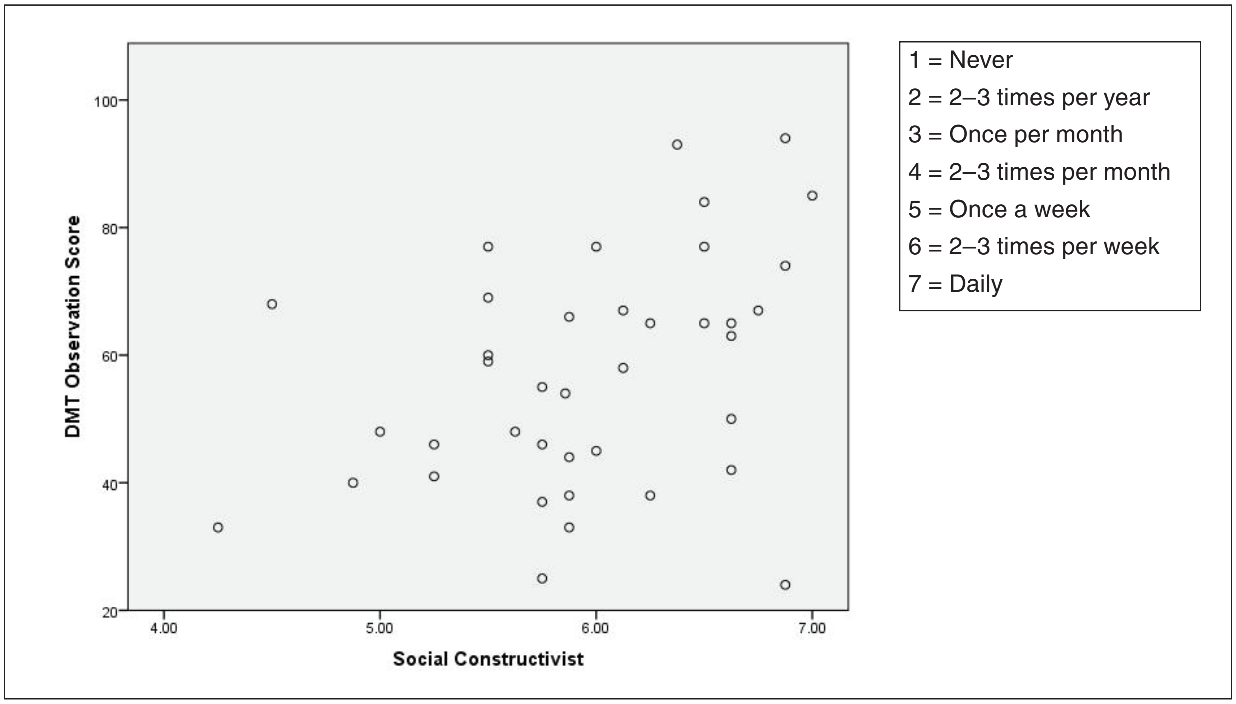

Figure G1. **Scatterplot of relationship between social-constructivist and the DMT observation instrument scores r = (37) .37, p < .05.**

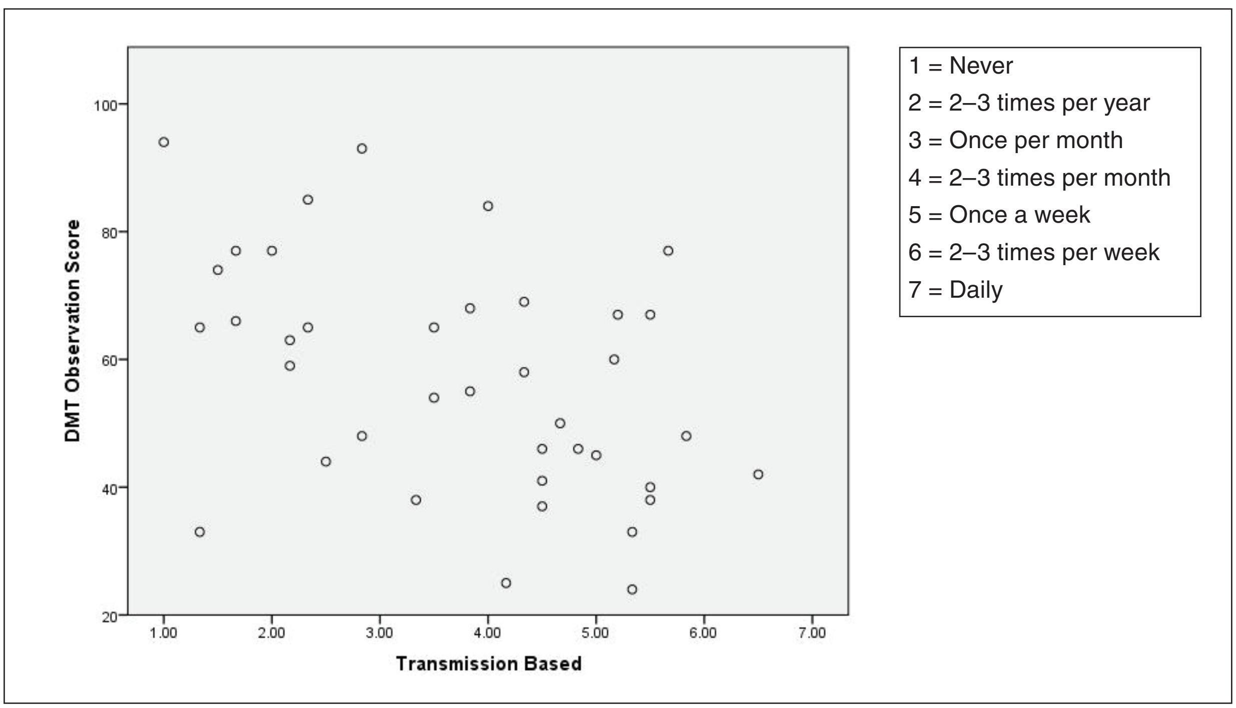

Figure G2. **Scatterplot of relationship between transmission and the DMT observation instrument scores r = (37) –.45, p < .01.**

(Return to page 104)