CrossMark

ORIGINAL ARTICLE

Using the Fennema-Sherman Mathematics Attitude Scales with lower-primary teachers

Lixin Ren[1] · **Jennifer L. Green**[2] · **Wendy M. Smith**[1]

Abstract The Fennema-Sherman Mathematics Attitude Scales (FSMAS) are among the most popular instruments used in studies of attitudes toward mathematics. However, the FSMAS has been mainly used among student populations and rarely used with teachers. In the present study, three scales from the FSMAS—*Confidence*, *Effectance Motivation*, and *Anxiety*—were revised and used with lower-primary (kindergarten to third grade) teachers. This study includes three parts: (1) a pilot study to ensure the modifications made to the FSMAS were appropriate to use with teachers, (2) confirmatory factor analyses to assess the factor structure of the revised FSMAS with 225 lower-primary teachers, and (3) measurement invariance analyses using data from a similar sample of 171 lower-primary teachers to examine whether the revised FSMAS measures each construct in the same way as in the previous sample. The final three-factor model, after removing three problematic items, achieves acceptable model fit, with each construct meeting all conditions for strict measurement invariance. Additionally, repeated measures analyses were performed on data collected from 39 in-service lower-primary teachers who participated in an elementary mathematics specialist program to examine the use of the revised FSMAS in program evaluation. Overall results suggest that researchers and program evaluators may use the revised FSMAS to reliably measure lower-primary teachers' mathematical attitudes, and it can be a valuable tool for evaluating the effectiveness of professional development programs.

✉ Lixin Ren
  lixin.ren@huskers.unl.edu

[1] Center for Science, Mathematics, & Computer Education, University of Nebraska—Lincoln, 251 Avery Hall, Lincoln, NE 68588-0131, USA

[2] Department of Mathematical Sciences, Montana State University, P.O. Box 172400, Bozeman, MT 59717-2400, USA

In this article, the research team describes how they adapted the Fennema-Sherman Mathematics Attitude Scales (FSMAS) to assess teachers' attitudes toward mathematics. The research team also discusses the rigorous procedures used to validate this measure's use with lower-primary teachers[1] (with a version of the FSMAS the team calls FSMAS-T) and the corresponding results of these analyses. Finally, the team provides an example of a study that used the FSMAS-T to examine longitudinal changes in teachers' mathematical attitudes after participating in an elementary mathematics specialist professional development program. While little research has been conducted to examine the longitudinal change in teachers' mathematical attitudes, the results of this study suggest that researchers and program evaluators may use the FSMAS-T to examine the impact of professional development on such changes (or stability).

The current study is situated within the educational context of a state and a number of districts within the USA. During the pilot study, the research team focused on answering the research question: Can the FSMAS be adapted for use with lower-primary teachers such that the adaptation forms valid and reliable scales for measuring teacher attitudes toward learning mathematics? Following that, the research team examined the research question: Does using the FSMAS-T with lower-primary teachers result in the same factor structure as the original FSMAS? Confirmatory factor analyses were conducted with data from lower-primary teachers to examine the psychometric qualities (e.g., dimensionality and reliability) of the adapted measure. Finally, the research team explored the following question: Does the FSMAS-T exhibit measurement invariance when used with lower-primary teachers? Measurement invariance analyses were conducted using data from a similar lower-primary teacher sample to cross-validate that the psychometric properties obtained from the previous step could be generalized across groups.

Literature review

A decade ago, Goldin (2002) expressed the concern that research in mathematics education was focused mainly on cognition and less on affect. Now, researchers recognize the importance of affective factors in explaining individual differences in the learning of mathematics (Philipp 2007). Affective factors influence one's decisions about how many mathematics courses one is willing to take, the amount of effort one is willing to exert in learning mathematics, and the way one approaches learning mathematical content (Reyes 1984).

Although the affective domain has been defined in various ways in the literature, McLeod (1992) proposed that attitudes, beliefs, and emotions were the major components of this domain in mathematics education. In the present study, the research team focuses on one of these components—attitudes. Specifically, the team examines lower-primary teachers' attitudes toward learning mathematics.[2] Neale (1969) defined attitudes toward mathematics as "a liking or disliking of mathematics, a tendency to

---

[1] In this study, the sample was composed of kindergarten to third grade teachers of students, ages 5 to 9 years, in the USA. For addressing an international audience, the research team refers to this level as "lower primary."

[2] The research team uses "mathematical attitudes" and "attitudes toward learning mathematics" interchangeably.

engage in or avoid mathematical activities, a belief that one is good or bad at mathematics, and a belief that mathematics is useful or useless" (p. 632). This definition refers to one's attitudes as a learner of mathematics and, although relatively old, captures the potential components of an individual's attitudes toward learning mathematics, such as confidence, motivation, and anxiety.

Within the affective domain, the distinction between attitudes, beliefs, and emotions remains unclear, and attitudes seem to overlap with beliefs, emotions or feelings, and values (Leder and Grootenboer 2005). For instance, mathematics anxiety has been characterized as fear (a "hot" emotion) or dislike (an attitude), and it may even include aspects that are cognitive (e.g., worry) (Ma 1999; McLeod 1992). In fact, McLeod (1992) conceptualized these components as a continuum:

> …we can think of beliefs, attitudes and emotions as representing increased levels of affective involvement, decreased levels of cognitive involvement, increasing levels of intensity of response, and decreasing levels of response stability (McLeod 1992, p. 579).

The research team acknowledges that individuals' confidence, motivation, and anxiety toward mathematics also may constitute other components of the affective domain. Therefore, in this article, mathematical attitudes refer to a broad construct of individuals' attitudes toward mathematics (Ma and Kishor 1997).

An individual's mathematical attitudes play an important role in the learning of mathematics. Many researchers have investigated students' attitudes toward learning mathematics and have consistently found that students' mathematical attitudes are related to their mathematics achievement and self-efficacy in mathematics (e.g., Cooper and Robinson 1991; Jackson 2015; Ma 1999; Muis and Foy 2010; Sherman and Christian 1999). Furthermore, Lipnevich et al. (2011) found that students' mathematical attitudes predicted their mathematics achievements independent of their innate mathematics ability. Individuals tend to perform more poorly than their actual abilities if they are anxious about mathematics, because anxiety and fears they experience when doing mathematics may prevent them from utilizing the mathematics knowledge they possess (Ashcraft and Kirk 2001). Negative feelings, such as worries and fears, can compromise the thinking and reasoning process (Beilock 2008) and can undermine self-confidence (Sherman and Christian 1999).

Less research has been focused on teachers' attitudes toward learning mathematics, particularly practicing, in-service teachers' mathematical attitudes and how their attitudes are related to their teaching practices and students' learning (McAnallen 2010). In one study, Wilkins (2010) investigated elementary school teachers' attitudes toward different subjects, and mathematics was consistently ranked as teachers' least favorite subject to teach, particularly for lower-primary teachers. Moreover, a large percentage of in-service teachers have reported anxiety toward mathematics (Haycock 2001), and teachers with high levels of mathematical anxiety tend to demonstrate low confidence in teaching elementary mathematics (Bursal and Paznokas 2006; Jackson 2015; Sherman and Christian 1999; Swars et al. 2006).

Teachers' mathematical anxiety can also be "contagious" to students' attitudes toward mathematics (Beilock et al. 2010; Stipek et al. 2001). Beilock et al. (2010) examined the relations between elementary female teachers' mathematical anxiety and

their students' mathematical attitudes and mathematics achievement. Teachers' mathematical anxiety was measured using the short Mathematics Anxiety Rating Scale (Alexander and Martray 1989), in which teachers responded to questions about how anxious different situations would make them feel (e.g., "studying for a math test"). The findings showed that at the beginning of the school year, there was no connection between female teachers' mathematical anxiety and their students' mathematics achievement. However, by the end of the school year, the more anxious the teachers were toward mathematics, the more likely girls (but not boys) in their classrooms were to believe the commonly held stereotype that "boys are good at math, and girls are good at reading." Moreover, girls who endorsed this stereotype had lower mathematics achievement than girls who did not and lower achievement than boys overall.

Elementary school mathematics experiences are often the first causes of negative mathematical attitudes, which in turn undermine students' confidence in their mathematical abilities and lead to mathematics avoidance later in school (Harper and Daane 1998). Thus, teachers' negative attitudes toward their own learning of mathematics can exert a significant impact on their students' current, as well as future, mathematics learning and achievement. In addition, with about 85 % of early mathematics teachers being female (Goldring et al. 2013), it is important to understand how to reduce teachers' mathematical anxiety and foster positive mathematical attitudes through professional development.

In-service teachers' attitudes toward learning mathematics not only influence the attitudes and achievement of their students but also their own instructional practices (e.g., Karp 1991; Quinn 1997; Richardson 1996; Stipek et al. 2001; Thompson 1992; Wilkins 2008). For example, Wilkins (2008) asked teachers to report their enjoyment and liking of mathematics, their perceptions of the importance of mathematics, their feelings of success with mathematics, and their enjoyment and liking of teaching mathematics. Findings revealed that teachers with more positive attitudes toward learning and teaching mathematics tended to have a stronger belief in the effectiveness of inquiry-based instruction (e.g., having students work in cooperative learning groups) and to use such instruction more frequently in their teaching. Inquiry-based instruction is often found to be positively related to student mathematics achievement (e.g., Hill et al. 2008).

Other studies also have explored relationships between teachers' mathematical attitudes and instructional practices. Karp (1991) found that teachers with negative attitudes toward mathematics more often used teacher-centered approaches and did not engage students actively in learning mathematics. Lee (2005) surveyed 200 Indiana kindergarten teachers and found that teachers' attitudes toward teaching mathematics were important predictors of the presence of developmentally appropriate teaching practices in the classroom. Stipek et al. (2001) conducted in-depth research of teachers of students aged 9–12 that involved the use of teacher surveys and multiple classroom observations. Stipek et al. found significant correlations between teacher attitudes and classroom practices as well as correlations between teachers' and students' self-confidence and enjoyment of mathematics. Therefore, teachers need to develop positive attitudes toward mathematics learning and teaching as well as their own feelings of efficacy as a mathematics learner and teacher (McLeod 1992; Pajares 1992).

With the increasing recognition of the importance of teachers' mathematical attitudes on their instructional practices and students' learning, it is crucial to explore ways

to improve teachers' mathematical attitudes, as well as to evaluate the effectiveness of those approaches. Teachers need to position themselves as life-long learners in order to become more effective in teaching, and the life-long learning process is an ongoing part of their professionalism (Graven 2004). Fennema and Franke (1992) proposed that instructional practice mediated the relationship between teacher characteristics (e.g., teachers' mathematical attitudes) and student learning. Thus, investigations of teachers' attitudes toward their own learning of mathematics will help develop a better understanding of teachers' learning and teaching, as well as potential learning opportunities that exist for students (Jong et al. 2012). Additionally, within the context of a teacher professional development program such as the one in which this study is situated, examining changes in teachers' attitudes toward learning mathematics is one potential indicator of program effectiveness.

To assess teachers' mathematical attitudes properly, sound measures are necessary. However, to the authors' knowledge, no reliable and validated measure of teachers' mathematical attitudes has been widely and consistently used in the literature. Several surveys exist that measure attitudes in mathematics, such as the Fennema-Sherman Mathematics Attitudes Scales (Fennema and Sherman 1976a), the Attitudes Toward Mathematics Inventory (Tapia and Marsh 2004), and the Mathematics Anxiety Rating Scale (Richardson and Suinn 1972). However, all three instruments were developed for use with students.

When planning the professional development program for lower-primary teachers in 2008, the research team was unable to find any measure of in-service teachers' attitudes toward learning mathematics that had been widely and consistently used in existing research. Since then, Welder et al. (2011) developed the Mathematics Experiences and Conceptions Surveys (MECS) to assess teachers' attitudes, beliefs, and dispositions toward the teaching and learning of mathematics. The MECS was developed and has been mainly used among preservice teachers (e.g., Hodges and Jong 2012; Welder and Jong 2012). The MECS includes both items assessing preservice teachers' attitudes toward teaching mathematics (e.g., "I look forward to teaching mathematics") and items pertaining to preservice teachers' attitudes toward learning mathematics (e.g., "I enjoy solving mathematics problems"). Even given the increased attention to teacher mathematical attitudes in the past decade, there still is no measure for in-service teachers' mathematical attitudes that is in wide use for research.

When measuring the impact of professional development on teacher participants, one often wants to measure changes in attitudes. Individuals' mathematical attitudes are shaped by their past experiences with mathematics, such as interactions with mathematics teachers in elementary school, teaching practices they experienced when they were in school, and experiences in past mathematics classes (e.g., Harper and Daane 1998; Jackson 2015; McAnallen 2010). Thus, frameworks focused on students' attitudes toward mathematics might be helpful in examining teachers' mathematical attitudes. Therefore, well-established measures that focused on students' mathematical attitudes might be informative in exploring teachers' mathematical attitudes.

The Fennema-Sherman Mathematics Attitudes Scales (Fennema and Sherman 1976a) are among the most popular instruments used in studies of students' attitudes toward mathematics. The FSMAS has been widely used by researchers in English-speaking countries besides the USA, including Canada (e.g., Chouinard et al. 2008) and Australia (e.g., Forgasz et al. 1999; Norton and Rennie 1998; Rowe 1993).

The FSMAS has also been translated into many languages, such as French (Vezeau et al. 1998), Chinese (Sachs and Leung 2007), and Arabic (Alkhateeb 2004). It has been used in international comparisons, such as in the Trends in International Mathematics and Science Study (TIMSS; Mullis et al. 2008).

Fennema-Sherman Mathematics Attitudes Scales

The Fennema-Sherman Mathematics Attitudes Scales (Fennema and Sherman 1976a) have been extensively used to assess students' attitudes toward learning mathematics. They were originally developed to capture gender-related differences in mathematical attitudes among high school students (Fennema and Sherman 1976a). The FSMAS encompasses nine scales, each of which measures domain-specific attitudes that are believed to be related to the learning of mathematics: *Attitude Toward the Success in Mathematics*, *Mathematics as a Male Domain*, *Mother* (perception of mother's attitudes toward one as a learner of mathematics), *Father* (perception of father's attitudes toward one as a learner of mathematics), *Teacher* (perception of teacher's attitudes toward one as a learner of mathematics), *Confidence in Learning Mathematics*, *Mathematics Anxiety*, *Effectance Motivation in Mathematics*, and *Mathematics Usefulness* (Fennema and Sherman 1976a). Each scale has 12 items, and responses are measured using a 5-point Likert scale (from "strongly disagree" to "strongly agree").

The nine scales can be used altogether or separately, and a review of the literature suggests that the FSMAS has been used widely and flexibly in research. Some researchers have used an individual scale in their research, while other researchers have used sets of two or more scales. For instance, Lim and Chapman (2013) and Dew et al. (1983) only used the Mathematics Anxiety scale in their studies, whereas Norton and Rennie (1998) used five scales (i.e., Mathematics Anxiety, Confidence in Learning Mathematics, Effectance Motivation, Attitude toward Success in Mathematics, and Mathematics as a Male Domain) of the FSMAS to examine students' mathematical attitudes in single-sex and coeducational schools. Other researchers have adapted one or more scales. For example, Betz (1978) adapted the Anxiety scale for use among college students, and Bai et al. (2009) made further revisions based on Betz's (1978) adapted version in their study. In other studies, researchers have selected items from the FSMAS instead of using a full scale. For instance, Pearn et al. (1996) selected items from the FSMAS and several other instruments to develop a multidimensional questionnaire to examine the mathematical attitudes of secondary students from schools in Australia with high percentages of students with non-English speaking backgrounds. Throughout these various uses, the FSMAS has been used primarily among student populations, particularly among secondary and college students, in many countries, and it has been rarely used among teachers.

Since the publication of the FSMAS, many researchers have re-examined its use among diverse populations. For instance, Forgasz et al. (1999) examined the Mathematics as a Male Domain scale among ninth-grade students from Australia, Sweden, and the USA and found that several items in the scale might no longer be valid and revisions were needed to ensure its usability in measuring the original construct. Lim and Chapman (2013) examined the factor structure of the Mathematics Anxiety scale using data from students attending pretertiary in Singapore (see Lim and Chapman

2013 for information on pretertiary institutions). Based on exploratory and confirmatory factor analyses, three items were removed and a two-factor model was retained (positively worded items formed one factor "FS-EASE" and negatively worded items made up the other factor "FS-ANX"). Melanchon et al. (1994)) investigated the measurement integrity of the FSMAS using data provided by public elementary school teachers in the USA, and the results supported the validity of the FSMAS for use with public elementary school teachers.

Shortened versions of the FSMAS also have been developed and used. For example, Mulhern and Rae (1998) developed a short form using principle component analysis, and Sachs and Leung (2007) developed a shortened version using Bhargava and Ishizuka's (1981) BI-method (a method that selects variables by using the trace information of the correlation or covariance matrix). However, the research team decided to use all items from a few scales instead of using five to six items from all the shortened scales. By using all items from the most teacher-relevant scales, the team hoped to capture a more in-depth picture of teachers' attitudes toward mathematics learning with respect to the scales chosen.

In the current study, the research team originally chose four scales from the FSMAS: (1) the *Confidence in Learning Mathematics Scale*, which measures individuals' confidence in their abilities to learn and to perform well in mathematics; (2) the *Mathematics Anxiety Scale*, which measures individuals' "feelings of anxiety, dread, nervousness, and associated bodily symptoms related to doing mathematics" (Fennema and Sherman 1976a, p. 326); (3) the *Effectance Motivation Scale in Mathematics*, which measures whether individuals enjoy and seek challenges regarding mathematics; and (4) the *Mathematics Usefulness Scale*, which measures individuals' beliefs about the usefulness of mathematics currently and in relationship to their future education and vocation. The research team chose these scales to align both with the goals of the teacher professional development program and the research literature about teacher factors that may be related to student outcomes. Teacher anxiety, in particular, has been shown to negatively impact student outcomes, particularly for female students (e.g., Beilock et al. 2010). The professional development program specifically sought to decrease teachers' anxiety toward mathematics while increasing teachers' motivation and confidence related to mathematics teaching and learning and helping teachers see mathematics as useful. Thus, these factors were well aligned to measure the impact of the professional development program on these desired dimensions.

Since the research team was focused on adult teachers learning mathematics and needed the survey to be as short as possible (given the other data collection demands on teachers' time), the team decided not to include the scales about *Mother*, *Father*, and *Teacher*. The items on these three scales are worded in ways that clearly apply to students living at home with parents, whose success in mathematics would be heavily influenced by their perception of parent and teacher expectations. Because Forgasz et al. (1999) found that the *Mathematics as a Male Domain* scale needed to be updated into more modern language and situations, the research team decided not to use this scale. Finally, while the *Attitude Toward the Success in Mathematics* scale would potentially have been interesting, some of the items would not pertain to teachers taking graduate courses in mathematics (e.g., "It would be really great to win a prize in mathematics"). The professional development goals included increasing teachers' mathematical knowledge for teaching, confidence, and motivation while decreasing

their anxiety toward mathematics. The professional development leaders also expressed an interest in measuring whether teachers grow in their recognition of the usefulness of mathematics, even though this was not one of the program goals. Seeing mathematics as useful is one indicator of a positive attitude toward mathematics teachers could be conveying to students. Thus, the scales chosen represent the best match to the project goals and leaders' interests and have since been selected for use in research on preservice elementary teachers' attitudes/beliefs (Tsao 2014).

To prepare for the pilot study, the research team first reviewed each item from the chosen scales and made edits as needed to ensure the items were applicable to present-day teachers, rather than students. The research team rewrote items using future tense to present tense. For example, the word "will" was excluded from the original item "I *will* use mathematics in many ways as an adult." The research team also reworded some items to make them culturally appropriate and easier to understand. For example, "I have a knack of flubbing up math" was changed to "I have a tendency to mess up math." One item from the Mathematics Usefulness Scale was eliminated (i.e., "I expect to have little use for mathematics when I get out of school"), because it is relevant for high school students, but not for teachers in general. The revised instrument was presented to a team of leading experts in the fields of mathematics education, early childhood education, mathematics, and statistics. The experts provided feedback on the relevance of items to lower-primary teachers, the extent to which lower-primary teachers might interpret the wording of the items in consistent ways, the alignment of the items to the overall goals of the professional development program, and the extent to which items might capture the intended change in teacher attitudes after participating in the professional development program.

Next, the research team randomized the order of the items and shared an initial draft of the instrument with a very experienced primary teacher leader. The teacher leader completed the draft survey and offered suggestions for further revisions. The research team also rewrote the item "Math has been my worst subject" to instead be "Math was my worst subject." Based on her feedback, the research team switched the order of some items. After making these changes, the instrument was ready to be piloted with a larger group of teachers.

Pilot study

The research team first conducted a pilot study to evaluate the FSMAS-T. Sixty-five teachers completed the FSMAS-T, including 43 elementary teacher leaders, two preservice teachers, and 20 middle-school teachers. The pilot teachers were involved in one of two different professional development programs, one held in a Southern state and one held in a Midwestern state. Each program was separately led by two of the experts who had helped review the instrument. Even though the research team did not collect teacher-level ethnicity data, the professional development leaders reported that the participants were nearly all Caucasian; thus, the research team knew they would not have enough teachers to look at subgroups by ethnicity. The 20 middle-school teachers were 90 % female. The elementary teacher leaders and preservice teachers were all female. Table 1 presents the characteristics for each of the samples involved in this

Table 1 Summary of samples

Sample	Size	% Male	Teaching assignment	Teaching experience, mean (SD)
Pilot study	65	3.08	Preservice = 3.08 % Primary school = 66.15 % Middle school = 33.85 %	NA
CFA sample	225	1.87	Grade K-3 = 84.68 % Multiple grades/grades 4–6 = 2.98 % Others = 3.83 % Unknown = 8.51 %	11.87 years (9.73)
Cross-validation sample	171	3.68	Grade K-3 = 84.21 % Multiple grades/grades 4–6 = 8.77 % ELL/special ed/coach = 2.92 % Others = 1.75 % Unknown = 2.34 %	9.26 years (6.35)
Evaluation study sample	39	2.56	Grade K-3 = 87.18 % Multiple grades/grades 4–6 = 7.69 % ELL/special ed/coach = 5.13 %	6.82 years (5.82)

Note. "Grade K-3" includes teachers teaching kindergarten to third grade; "multiple grades/grades 4–6" includes teachers teaching multiple grades or fourth to sixth grade; "ELL/special ed/coach" includes those who were English language learners teachers, special education teachers, or math coaches; "others" includes media center teachers, integrated technology specialists, and other teaching roles

study, including the percentage of male teachers in the sample, years of teaching experience, and teaching assignment at the time of the study.

For each predefined scale, the research team conducted reliability analyses on the pilot data to identify potentially problematic items. Based on the coefficient alphas, the *Confidence in Learning Mathematics* (Confidence) ($\alpha = 0.906$), *Mathematics Anxiety* (Anxiety) ($\alpha = 0.924$), *Effectance Motivation in Mathematics* (Effective Motivation) ($\alpha = 0.921$), and *Mathematics Usefulness* (Usefulness) ($\alpha = 0.858$) scales had acceptable estimated reliability for the teachers. For each scale, the research team also examined items with low item-total correlations. Two items from the Confidence scale and one item from each of the other scales had item-total correlations less than 0.4. However, none of these items were edited or removed from the final instrument, because the items measured different, but important, aspects of their respective constructs.

Descriptive statistics also were obtained for each item. Responses for six of the items on the Usefulness scale had low variability. Specifically, most of the pilot participants responded positively to these items ("agree" or "strongly agree"), indicating that most of the teachers perceived mathematics as useful. The research team wanted to be able to detect potential positive changes in teachers' mathematical attitudes after participating in professional development programs, and thus, needed scales with room for growth. Additionally, movement on this scale was not deemed an important objective of the larger research study around program impact. No program goals were related to teachers finding mathematics more useful, but improved attitudes, including decreased anxiety and increased confidence and motivation, were explicit goals of the professional development. Therefore, the Usefulness scale was eliminated from the final

survey (see Appendix for the survey; items are arranged by scale). Advanced analysis, such as confirmatory factor analysis, was not conducted because of the small sample size. The teachers also had the opportunity to provide qualitative feedback about the instrument and any item(s) they found difficult to understand and/or answer. However, none of the written responses indicated issues with the individual items that needed to be addressed.

Instrument validation

Results of the pilot study suggest that the FSMAS can be adapted for use with teachers. However, due to the small sample size, only basic descriptive and reliability analyses were conducted. To further validate the use of the FSMAS-T with teachers, a rigorous validation procedure was performed. The validation procedure for the FSMAS-T involved two steps. First, a series of confirmatory factor analyses were conducted with data from a sample primarily comprised of lower-primary teachers (the confirmatory factor analysis sample) to examine the factor structure of the FSMAS-T. Confirmatory factor analysis is a theory-driven statistical procedure that is used to verify the factor structure of a set of observed variables (e.g., measurement items). Next, to ensure that the factor structure obtained from the first step was not sample-specific, measurement invariance analyses were conducted using data from a similar sample that primarily consisted of lower-primary teachers (the cross-validation sample) to examine whether the psychometric properties of the FSMAS-T obtained from the confirmatory factor analysis sample could be generalized to this group. Measurement invariance (i.e., measurement equivalence) analysis concerns the extent to which the psychometric properties of the observed indicators can be generalized across groups or over time/condition.

Confirmatory factor analysis sample

The confirmatory factor analysis sample came from a larger National Science Foundation-funded Math Science Partnership study examining the effectiveness of a professional development program. A total of 225 teachers completed the FSMAS-T, and they were followed for several years to examine longitudinal changes in their mathematical attitudes. Participating districts included four larger core partner districts and 28 smaller districts across Nebraska, a Midwestern state in the USA. About 85 % of the teachers were teaching kindergarten to third grade at the time of the study. About 3 % of the teachers were teaching upper-primary grades (fourth to six grade) or multiple grades, and 4 % of the teachers had other teaching assignments. Almost 98 % of the participants were female teachers, and they were primarily Caucasian. The teachers had an average of 12 years of teaching experience, and all of the teachers had a bachelor's degree or higher.

In this study, teachers' baseline data (the first time they completed the FSMAS-T) were used to conduct confirmatory factor analyses, because these data are more representative of teacher attitudes toward mathematics prior to participating in professional development. The baseline data also provide a basis for which to measure change.

Cross-validation sample

The cross-validation sample is composed of 171 teachers who came from most of the same districts and had similar characteristics as the confirmatory factor analysis sample. The original Math Science Partnership research study included teachers participating in the professional development during 2009–2011. As the program transitioned to an institutionalization phase, additional groups of teachers participated in the professional development program in 2012–2014. These teachers also were followed for several years, and they completed the FSMAS-T multiple times. In this study, only teachers' baseline data were used in the measurement invariance analyses. At baseline, they had an average of 9 years of teaching experience. About 84 % of the teachers were teaching lower-primary grades (kindergarten to third grade); 9 % of the participants were upper-primary teachers (fourth to sixth grade) or teaching multiple grades; and 3 % of them were English language learner teachers, special education teachers, or math coaches. In addition, only 4 % of the teachers were male. Thus, similar to the confirmatory factor analysis (CFA) sample, the cross-validation sample primarily was composed of female Caucasians.

Results

Confirmatory factor analysis

Robust maximum likelihood estimation in Mplus v. 7.11 (Muthén and Muthén 1998–2013) was used in confirmatory factor analyses to examine the psychometric qualities of the FSMAS-T. The research team did not perform an exploratory factor analysis. One main goal of exploratory factor analysis is to reduce a set of variables to a smaller number of variables while retaining as much of the original variance as possible. In this case, the use of exploratory factor analysis is more pragmatic than theoretical (Brown 2006). Based on Fennema and Sherman's (1976a) findings, each scale in the FSMAS measured a single construct. The research team did not intend to reduce the number of variables (i.e., items), but to test whether each scale would measure the same construct as the original FSMAS for a different population. Thus, as a theory-driven analytical approach, confirmatory factor analysis is appropriate for this purpose.

First, the research team conducted a confirmatory factor analysis using the 225 teachers' baseline data to examine the construct of each scale. Consistent with the original FSMAS, the research team proposed that each scale is unidimensional, which means that each item is constrained to measure only one construct. Thus, to test the unidimensionality of the three factors—namely *Confidence*, *Effectance Motivation*, and *Anxiety*—the team conducted a confirmatory factor analysis for each scale separately. Table 2 presents the model fit statistics of all the CFA models examined. Figure 1 presents the final measurement model.

The research team used several model fit indices to evaluate the model fit, including the obtained model chi-square value (χ^2), the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Although there are other fit indices available, the research team only used the ones listed

Table 2 Model fit indices of the confirmatory factor analyses

| Model	No. of items	χ^2	$\chi^2 p$ value	CFI	TLI	RMSEA	SRMR
Confidence (full scale)	12	190.246	<0.001	0.896	0.873	0.106	0.057
Confidence (item 5 removed)	11	144.969	<0.001	0.912	0.890	0.101	0.054
Confidence (item 16 removed)	11	119.510	<0.001	0.934	0.918	0.087	0.049
Confidence (items 5 and 16 removed)	10	89.609	<0.001	0.945	0.929	0.083	0.046
Effectance motivation (full scale)	12	123.513	<0.001	0.941	0.928	0.076	0.041
Effectance motivation (item 1 removed)	11	80.719	<0.001	0.964	0.955	0.061	0.036
Effectance motivation (items 1 and 26 removed)	10	67.870	<0.001	0.964	0.953	0.065	0.037
Anxiety (full scale)	12	130.842	<0.001	0.943	0.931	0.080	0.038
Three-factor model	33	897.771	<0.001	0.911	0.905	0.060	0.054
One-factor model	33	1227.542	<0.001	0.840	0.829	0.081	0.131
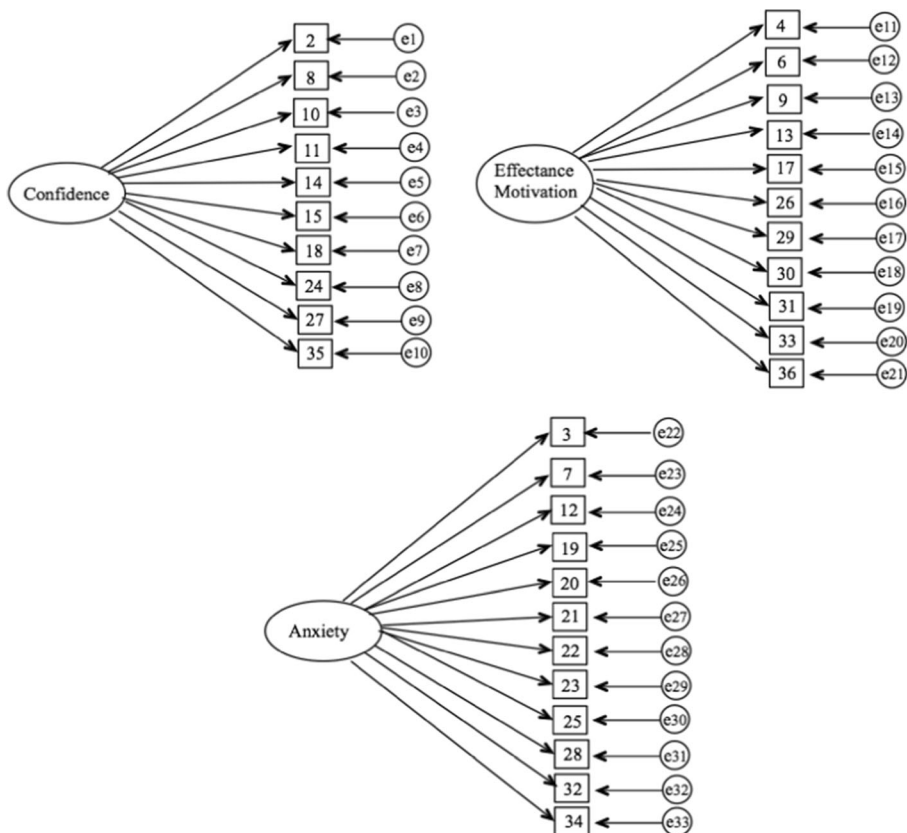
Fig. 1 The final measurement model of the FSMAS-T after removing three problematic items. The model fit of each factor was examined separately in the analysis, and the standardized and unstandardized model estimates can be found in Table 3

above based on Brown's (2006) recommendation and findings from a study by Jackson et al. (2009). Brown (2006) selected these indices on the basis of their popularity in the literature, as well as their favorable performance in simulation studies. After examining almost 200 studies, Jackson et al. (2009) also found that chi-square values, CFI, RMSEA, and TLI are the most commonly used fit indices. In general, using these different types of model fit indices in conjunction with one another provides stronger evidence for model fit than solely relying on one model fit index (Brown 2006).

For each fit index, the research team used the criteria reported by Brown (2006) to evaluate model fit. For example, the research team used the obtained model χ^2 to assess model fit at an absolute level. A significant χ^2 test ($p < 0.05$) indicates that the exact model fit is not good, and nonsignificance is desirable. However, because χ^2 is susceptible to sample size, the research team also used other model fit indices, as is customary. For CFI and TLI, values greater than 0.95 indicate good fit, and values greater than 0.9 indicate acceptable model fit. For RMSEA, values smaller than 0.06 are desirable for good fit, values of 0.06 to 0.08 indicate acceptable model fit, and values of 0.08 to 0.10 indicate mediocre fit. The 90 % confidence interval of RMSEA is also reported. SRMR values of 0.08 or smaller indicate good model fit. Together, these criteria helped the research team evaluate the fit of each CFA model. In addition to the various fit indices, the research team examined standardized factor loadings and squared multiple correlations (R^2) of the items to investigate effect size. A factor loading represents the correlation between the item and the latent factor; R^2 represents the amount of variance in individuals' responses to the item that can be attributed to the latent factor. The research team also examined normalized residuals for diagnosing local misfit to respecify the model when needed.

For the *Confidence* scale, the model fit was not acceptable, $\chi^2(54) = 190.246$ ($p < 0.001$), CFI = 0.896, TLI = 0.873, RMSEA = 0.106 (90 % CI = 0.090~0.122), SRMR = 0.057. The research team inspected the normalized residual matrix and discovered that two items were more correlated than the model indicated (i.e., item 5: "I don't think I could do advanced mathematics"; item 16: "I am sure I could do advanced work in mathematics"). The extra correlation between the two items might be due to the common word "advanced." The phrase "advanced" might have caused problems, as lower-primary teachers may interpret "advanced" as courses like calculus, which they may see as irrelevant to teaching lower-primary mathematics. To retain as many items as possible, the research team examined model fit while removing only one item at a time. Removing item 5 did not result in desirable model fit, especially according to RMSEA, $\chi^2(44) = 144.969$ ($p < 0.001$), CFI = 0.912, TLI = 0.890, RMSEA = 0.101 (90 % CI = 0.083~0.119), SRMR = 0.054. Removing item 16 led to acceptable model fit according to CFI, TLI, SRMR, but not χ^2 or RMSEA, $\chi^2(44) = 119.510$ ($p < 0.001$), CFI = 0.934, TLI = 0.918, RMSEA = 0.087 (90 % CI = 0.069~0.106), SRMR = 0.049. Thus, the research team decided to remove both items containing the word "advanced," which resulted in a better model fit, $\chi^2(35) = 89.609$ ($p < 0.001$), CFI = 0.945, TLI = 0.929, RMSEA = 0.083 (90 % CI = 0.062~0.105), SRMR = 0.046. All the items had significant standardized factor loadings ranging from 0.496 to 0.870 and R^2 values that ranged from 0.246 to 0.757.

Next, composite reliability (i.e., coefficient Omega) was calculated. Composite reliability is conceptually similar to the internal consistency coefficient—coefficient alpha. However, the calculation of coefficient alpha makes the assumption that all items load on a single underlying construct and all items represent that construct equally well (i.e., equal factor loadings). In confirmatory factor analysis, indicators (i.e., items) are allowed to have heterogeneous correlations with the underlying factor (i.e., heterogeneous factor loadings). Thus, composite reliability provides a more precise estimate of reliability than coefficient alpha (Geldhof et al. 2014). In the formula for calculating coefficient Omega, the numerator is the squared sum of factor loadings (i.e., the true score variance), and the denominator represents the true score variance plus all residual variances. The *Confidence* scale obtained a composite reliability coefficient of 0.935.

For the *Effectance Motivation* scale, the one-factor model fit was acceptable according to most fit indices, $\chi^2(54) = 123.513$ ($p < 0.001$), CFI = 0.941, TLI = 0.928, RMSEA = 0.076 (90 % CI = 0.058~0.093), SRMR = 0.041. When inspecting the normalized residual matrix, the research team found that item 1 "I like math puzzles" might be problematic. The normalized residual matrix indicated this item should be less correlated with other items than what the model estimated, suggesting that this item may not measure the same construct as the other items. The research team also found this item to be problematic in the pilot study and hypothesized that different people might have different understandings regarding what a math puzzle is. While there is another item that uses the phrase "math puzzle" (i.e., item 26: "Math puzzles are boring"), that item did fit in the model and seemed to get more at teachers' interests, whereas the former (i.e., "I like math puzzles") may imply some actions—that a teacher would seek out math puzzles due to liking them.

Because the research team wanted to retain as many items as possible to better measure the construct, the team first removed only one item, "I like math puzzles," which resulted in improved model fit, $\chi^2(44) = 80.719$ ($p < 0.001$), CFI = 0.964, TLI = 0.955, RMSEA = 0.061 (90 % CI = 0.039 ~ 0.082), SRMR = 0.036. All the items had significant standardized factor loadings ranging from 0.388 to 0.878, and R^2 values ranged from 0.151 to 0.771. Removing both items 1 and 26 that contain the phrase "math puzzles" resulted in slightly worse model fit than the model with only item 1 removed, $\chi^2(35) = 67.870$ ($p < 0.001$), CFI = 0.964, TLI = 0.953, RMSEA = 0.065 (90 % CI = 0.041~0.087), SRMR = 0.037. This evidence, along with the research team's desire to retain as many items as possible, led the team to decide not to remove item 26, "Math puzzles are boring." Furthermore, this item fit the initial model, as indicated by the normalized residual matrix discussed previously. The composite reliability coefficient for the 11-item *Effectance Motivation* scale was 0.918.

For the *Anxiety* scale, the model exhibited an acceptable fit, $\chi^2(54) = 130.842$ ($p < 0.001$), CFI = 0.943, TLI = 0.931, RMSEA = 0.080 (90 % CI = 0.062~0.097), SRMR = 0.038. All of the items had significant factor loadings that ranged from 0.540 to 0.903 and significant R^2 values that ranged from 0.292 to 0.815. The *Anxiety* scale had a composite reliability coefficient of 0.942. Thus, in the final models, the research team eliminated two items from the *Confidence* scale, removed one item from the *Effectance Motivation* scale, and retained all items from the *Anxiety*

Table 3 Model estimates of the confidence, effectance motivation, and anxiety factors in the final models after removing problematic items

Item	Unstandardized		Standardized		
	Loading (standard error)	Intercept (standard error)	Loading (standard error)	Intercept (standard error)	R^2 (standard error)
Confidence					
Item 2	0.727 (0.060)	3.489 (0.066)	0.735 (0.041)	3.527 (0.190)	0.540 (0.061)
Item 8	0.319 (0.045)	4.240 (0.043)	0.496 (0.060)	6.594 (0.420)	0.246 (0.059)
Item 10	0.825 (0.050)	3.530 (0.065)	0.840 (0.026)	3.596 (0.194)	0.706 (0.043)
Item 11	0.746 (0.057)	3.434 (0.068)	0.736 (0.043)	3.391 (0.169)	0.542 (0.063)
Item 14	1.073 (0.056)	3.436 (0.083)	0.861 (0.031)	2.757 (0.135)	0.742 (0.053)
Item 15	0.850 (0.053)	2.960 (0.070)	0.812 (0.037)	2.827 (0.115)	0.659 (0.060)
Item 18	0.607 (0.056)	3.351 (0.060)	0.672 (0.047)	3.707 (0.189)	0.452 (0.063)
Item 24	0.688 (0.054)	3.775 (0.057)	0.805 (0.032)	4.415 (0.275)	0.648 (0.052)
Item 27	0.453 (0.054)	3.878 (0.047)	0.643 (0.054)	5.501 (0.389)	0.414 (0.069)
Item 35	0.840 (0.043)	3.403 (0.064)	0.870 (0.020)	3.524 (0.170)	0.757 (0.035)
Effectance motivation					
Item 4	0.522 (0.062)	3.717 (0.058)	0.602 (0.050)	4.288 (0.272)	0.362 (0.060)
Item 6	0.833 (0.051)	3.560 (0.068)	0.811 (0.028)	3.465 (0.174)	0.658 (0.045)
Item 9	0.578 (0.057)	3.508 (0.060)	0.641 (0.048)	3.890 (0.215)	0.411 (0.061)
Item 13	0.355 (0.079)	3.551 (0.061)	0.388 (0.077)	3.888 (0.248)	0.151 (0.060)
Item 17	0.787 (0.058)	3.437 (0.065)	0.812 (0.045)	3.545 (0.195)	0.659 (0.074)
Item 26	0.493 (0.052)	3.874 (0.049)	0.669 (0.049)	5.263 (0.316)	0.448 (0.066)
Item 29	0.825 (0.051)	3.375 (0.070)	0.793 (0.026)	3.243 (0.164)	0.629 (0.041)
Item 30	0.835 (0.050)	3.483 (0.064)	0.878 (0.023)	3.660 (0.200)	0.771 (0.040)
Item 31	0.670 (0.051)	3.284 (0.063)	0.712 (0.039)	3.489 (0.160)	0.506 (0.055)
Item 33	0.764 (0.050)	3.396 (0.067)	0.768 (0.031)	3.416 (0.166)	0.590 (0.048)
Item 36	0.505 (0.062)	3.714 (0.055)	0.609 (0.059)	4.485 (0.259)	0.371 (0.071)
Anxiety					
Item 3	0.656 (0.062)	2.676 (0.068)	0.642 (0.054)	2.621 (0.095)	0.412 (0.070)
Item 7	0.631 (0.057)	3.684 (0.062)	0.682 (0.047)	3.982 (0.236)	0.465 (0.064)
Item 12	0.834 (0.047)	2.982 (0.069)	0.804 (0.030)	2.873 (0.114)	0.646 (0.048)
Item 19	0.789 (0.058)	3.486 (0.068)	0.781 (0.033)	3.449 (0.200)	0.610 (0.051)
Item 20	0.500 (0.066)	3.194 (0.062)	0.540 (0.066)	3.453 (0.167)	0.292 (0.071)
Item 21	0.814 (0.057)	3.292 (0.066)	0.819 (0.047)	3.315 (0.161)	0.671 (0.077)
Item 22	0.726 (0.072)	3.398 (0.069)	0.700 (0.059)	3.278 (0.157)	0.491 (0.083)
Item 23	0.884 (0.050)	3.310 (0.070)	0.846 (0.026)	3.168 (0.154)	0.716 (0.044)
Item 25	0.919 (0.047)	3.435 (0.068)	0.903 (0.018)	3.376 (0.171)	0.815 (0.033)
Item 28	0.620 (0.055)	3.563 (0.064)	0.654 (0.042)	3.758 (0.200)	0.427 (0.055)
Item 32	0.826 (0.043)	3.485 (0.062)	0.886 (0.019)	3.738 (0.185)	0.785 (0.033)
Item 34	0.737 (0.050)	2.947 (0.064)	0.774 (0.033)	3.096 (0.118)	0.599 (0.051)

Note. The *p* values for all the estimates of the unstandardized and standardized factor loadings, intercepts, and R^2 values are all smaller than 0.001

scale. Table 3 provides the estimates and standard errors for the item factor loadings from the final standardized solutions for all three factors.

Finally, a three-factor model was examined. In this model, the *Confidence*, *Effectance Motivation*, and *Anxiety* factors were allowed to be correlated. The model obtained an acceptable fit, $\chi^2(492) = 896.771$ ($p < 0.001$), CFI = 0.911, TLI = 0.905, RMSEA = 0.060 (90 % CI = 0.054~0.067), SRMR = 0.054. The three factors were highly correlated, with correlation coefficients ranging from 0.837 to 0.962. Thus, a one-factor model with all 33 remaining items was examined to investigate whether there was actually only one factor instead of three factors. This model did not have an acceptable fit, $\chi^2(496) = 1227.542$ ($p < 0.001$), CFI = 0.840, TLI = 0.829, RMSEA = 0.081 (90 % CI = 0.075~0.087), SRMR = 0.131. Therefore, the research team recommends retaining the three scales instead of combining the three scales into one scale. However, high correlations among the three scales should be expected, which is consistent with previous findings (Fennema and Sherman 1976b; Reyes 1984).

Measurement invariance analysis

To ensure the final factor structures obtained from the previous analyses are not specific to the studied sample, measurement invariance analyses were conducted using data from the confirmatory factor analysis sample and the cross-validation sample. Achieving measurement invariance means the same construct is measured in the same way between the two studied samples.

Testing measurement invariance includes a few steps. First, the research team used a "configural" invariance model as the baseline model against which all following models were tested. In the configural model, all factor loadings, intercepts, and residuals were estimated freely for the two samples. Thus, the research team thinks of this model as the "best" model with the most parameters. In subsequent steps, the research team used deviance tests (i.e., likelihood-ratio tests) to compare nested models in which additional modeling constraints were placed on factor loadings, intercepts, and residuals. The research team used robust maximum likelihood estimation and adjusted the deviance tests accordingly to account for the scaling factor.

In the second step, the research team tested "metric" invariance to determine whether the two samples had equal factor loadings. For each item, factor loadings were constrained to be the same across the two samples. Specifically, loadings for the same item were constrained to be equal between the CFA sample and the cross-validation sample. Meanwhile, each item was still allowed to have a different loading. This model was nested within the configural model. If the metric invariance model fits the same statistically as the configural model, it indicates the loadings are equal across groups. In this study, metric invariance was achieved for all three factors: *Confidence*, χ^2 difference (9) = 14.5, $p = 0.23$; *Effectance Motivation*, χ^2 difference (10) = 18.7, $p = 0.08$; and *Anxiety*, χ^2 difference (11) = 12.8, $p = 0.39$. Thus, the two samples had equal factor loadings for each common item.

In the third step, the research team tested "scalar" invariance to determine whether the two samples had equal item intercepts for the same item. In this

model, intercepts for the same item were constrained to be equal across the samples, as were the factor loadings. The scalar invariance model was nested within the metric invariance model. The deviance test showed that full scalar invariance was achieved for all three factors: *Confidence*, χ^2 difference $(9) = 8.0$, $p = 0.55$; *Effectance Motivation*, χ^2 difference $(10) = 6.7$, $p = 0.75$; and *Anxiety*, χ^2 difference $(11) = 13.7$, $p = 0.24$. Thus, in addition to equal factor loadings, the two samples had equal intercepts for each common item.

The final step assessed "residual" invariance in which residual variances for the same item were constrained to be equal across the two samples. The residual variance invariance model was nested within the scalar invariance model. For each item, the factor loadings and intercepts continued to be constrained to be equal across the samples. The deviance test showed that full residual variance invariance was achieved for all the factors: *Confidence*, χ^2 difference $(10) = 23.7$, $p = 0.30$; *Effectance Motivation*, χ^2 difference $(11) = 17.9$, $p = 0.30$; and *Anxiety*, χ^2 difference $(12) = 15.6$, $p = 0.60$. Thus, in addition to equal factor loadings and intercepts, the two samples had equal residual variances for each common item. To conclude, achieving measurement invariance suggests that the final constructs of the FSMAS-T were not specific to the confirmatory factor analysis sample. By cross-validating the constructs of the FSMAS-T with a similar (but separate) sample of lower-primary teachers, the research team has faith in the usability and reliability of the measurement.

Using the FSMAS-T to evaluate program effect

The research team used the FSMAS-T to measure the impact of a research-based, intensive professional development program, Primarily Math, in changing teachers' mathematical attitudes. Primarily Math is a six-course (18 credit hours), 13-month elementary mathematics specialist (EMS) program aimed to strengthen kindergarten to third grade teachers' pedagogical and mathematical content knowledge and to improve instruction to the extent that it creates measurable gains in mathematics achievements of kindergarten to third grade students across the state. Of the six courses, three are focused on increasing teachers' mathematical knowledge for teaching, while the other three are focused on improving teachers' knowledge in pedagogy and child development (refer to http://scimath.unl.edu/primarilymath/ and Kutaka et al. [under review] for detailed information regarding the sequence, content, and sample assignments of the six courses).

The Primarily Math leadership team designed the professional development program to help teachers acquire more knowledge in mathematical content, mathematics teaching, pedagogy, and child development. Through this program, it was expected that teachers would become more confident in their abilities to learn mathematics, more motivated to learn more mathematics, and less anxious toward learning mathematics. In Kutaka et al. (under review), the research team reported findings of the impact of Primarily Math on teachers' mathematical attitudes measured by the FSMAS-T; the teachers under investigation were a subset of teachers from the CFA sample. The findings showed that teachers who participated in Primarily Math had significant increases in their confidence and effectance

motivation and a significant decrease in their anxiety from pretest (the testing occasion prior to starting Primarily Math coursework) to posttest (the testing occasion immediately after the completion of Primarily Math). In addition, Primarily Math teachers showed larger increases in their confidence and effectance motivation, as well as a larger decrease in their anxiety toward learning mathematics relative to the comparison teachers whose scores remained statistically unchanged across the research study. Detailed results and discussion of the findings were described in Kutaka et al. (under review).

The Primarily Math research team also used the FSMAS-T to assess changes in teachers' attitudes toward learning mathematics among a group of 39 in-service teachers who participated in the Primarily Math professional development program during 2014–2015. These 39 teachers were not part of the CFA sample or the cross-validation sample, but they all came from one of the same core partner districts in which teachers in the CFA sample and the cross-validation sample taught. This district decided to use Title I[3] professional development money to fund 39 teachers to complete Primarily Math. There was an open application process, which followed the same procedures as those used to select the Primarily Math participants in the original research studies: the district chose 39 teachers (1 male) from over 120 who applied. The teachers had an average of 7 years of teaching experience. About 87 % of the teachers were teaching lower-primary grades (kindergarten to third grade); 8 % of them were teaching multiple grades or upper-primary grades (fourth to sixth grade), and 5 % of the teachers were special education teachers, English language learner teachers, or math coaches.

Throughout their participation in Primarily Math, teachers completed the FSMAS-T during four different measurement occasions. The first measurement occasion (pretest) was completed at the beginning of the summer of 2014, prior to starting any coursework. The second measurement occasion (midtest 1) was taken at the end of the summer of 2014, after completing two mathematics content courses. During the 2014–2015 academic year, teachers completed one pedagogy course each semester. The third measurement occasion (midtest 2) was completed at the beginning of the summer of 2015. During the summer of 2015, teachers completed the last two courses of Primarily Math, including one mathematics content course and one pedagogy course. The fourth measurement occasion (posttest) was taken in summer 2015, after the completion of all Primarily Math courses.

It was hypothesized that teachers would become more confident and motivated and less anxious toward mathematics after learning more about mathematics, mathematics teaching, and child development through Primarily Math. Mean confidence, effectance motivation, and anxiety scores were calculated. The research team fit a separate repeated measures ANOVA model for each of the three outcomes, in which the confidence, effectance motivation, or anxiety score was included as the dependent variable, and measurement occasion was specified as the predictor. For teacher-reported confidence, the effect of measurement occasion

---

[3] Title I is a federal program that gives school districts additional federal funds based on the percentage of students in poverty served by the district.
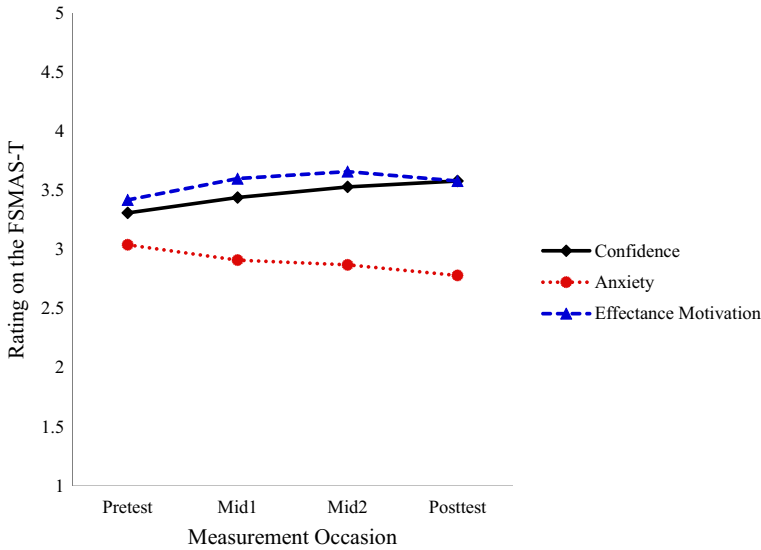
Fig. 2 Changes of Primarily Math teachers' attitudes toward learning mathematics

was significant, F (3, 108) = 6.42, $p < 0.001$. Post hoc comparisons showed teachers' confidence gradually increased; when compared to pretest scores, teachers reported more confidence at midtest 1 ($t = 2.03$, $p = 0.045$), midtest 2 ($t = 3.38$, $p = 0.001$), and posttest ($t = 4.07$, $p < 0.001$). Teacher-reported effectance motivation also changed significantly across measurement occasions, F (3, 108) = 3.70, $p = 0.014$. Post hoc comparisons indicated that compared to pretest scores, teachers reported higher levels of effectance motivation at midtest 1 ($t = 2.67$, $p = 0.001$), midtest 2 ($t = 3.05$, $p = 0.003$), and posttest ($t = 1.98$, $p = 0.050$). Finally, teacher anxiety also changed across time, F (3, 108) = 4.78, $p = 0.004$. Teacher anxiety gradually decreased over time; when compared to pretest scores, teachers reported lower levels of anxiety at midtest 2 ($t = 2.40$, $p = 0.018$) and posttest ($t = 3.73$, $p < 0.001$). Figure 2 presents teachers' confidence, effectance motivation, and anxiety at each of the four measurement occasions. Because there was not a comparison group for this set of participants, the research team cannot attribute these changes to participation in Primarily Math. Nevertheless, the findings reported by Kutaka et al. (under review), in conjunction with the results described above, provide evidence that the FSMAS-T is a valuable tool for evaluating the effectiveness of professional development in changing lower-primary teachers' attitudes toward learning mathematics.

Conclusions

The goal of this study was to determine if the research team could use the *Confidence*, *Effectance Motivation*, and *Anxiety* scales from the FSMAS (Fennema and Sherman 1976a) to assess lower-primary teachers' attitudes toward learning mathematics, and if so, to use the measure to assess the impact of a

professional development program. After using the pilot study results to choose which revised FSMAS scales to use, the research team was able to establish the instrument's reliability for use with teachers. Confirmatory factor analyses were conducted to investigate the factor structure of the revised measure. The research team found that each scale measured one latent construct after deleting a few problematic items. Measurement invariance analyses further validated the psychometric properties of the revised instrument.

The current study adds knowledge to the research of teachers' attitudes toward learning mathematics. Previous research suggests that teachers' mathematical attitudes influence students' attitudes toward mathematics, students' mathematics achievement, and teachers' instructional practices in the classroom (Beilock et al. 2010; Karp 1991; Stipek et al. 2001; Wilkins 2008). Thus, reliable and validated assessments of teachers' mathematical attitudes are needed. The FSMAS-T is useful for research investigating lower-primary teachers' attitudes toward learning mathematics, including research to study whether professional developmental programs have a direct impact on teachers' attitudes toward learning mathematics. Results from this instrument may increase researchers' and educators' understanding of teachers' mathematical attitudes, particularly with respect to their relationship with and influence on students' attitudes and mathematics achievement, as well as their impact on teachers' instructional practices in the classroom.

In this study, teacher samples came from a Midwestern state in the USA, and all but a few of the teachers are female. In addition, many of the teachers in the samples were enrolled in voluntary professional development. These characteristics aligned with the research team's target population, because the research team intended to use the FSMAS-T with lower-primary teachers participating in professional development, and worldwide, most countries have more female lower-primary teachers than male (The World Bank 2015). However, these features could limit the generalizability of the findings. Thus, while the research team was able to validate the FSMAS-T with the heavily female population of Midwest primary teachers in the USA, further research is necessary to validate the use of the instrument with male primary teachers, primary teachers in other areas of the USA, and primary teachers in other countries. The research team knows that the status of teaching as a profession has wide variation across different countries with different cultural contexts. Teacher credential requirements also vary widely by country (e.g., Tatto 2013), which can further result in wide variation in teachers' attitudes toward mathematics. Because the FSMAS has been used with students from many cultures (e.g., Alkhateeb 2004; Forgasz et al. 1999; Norton and Rennie 1998; Sachs and Leung 2007; Vezeau et al. 1998), it may also be appropriate to use the FSMAS-T with teachers from other cultures. In addition, the FSMAS-T may be applicable to preservice teachers and/or in-service mathematics teachers who teach upper-elementary grades, middle school, or secondary school. Additional research is needed to examine the use of the FSMAS-T among these populations. Once the usability of the FSMAS-T is established, the FSMAS-T can be used to compare teachers' mathematical attitudes across grade levels and across cultures.

Appendix

Table 4 The Fennema-Sherman Mathematics Attitude Scales—Teacher Form (FSMAS-T)

Item	Item description	Reverse coding	Keep/remove
Confidence in Learning Mathematics Scale:			
2	Most subjects I can handle O.K., but I have a tendency to mess up math.	Reverse	Keep
5	I don't think I could do advanced mathematics.	Reverse	Remove
8	I am sure that I can learn mathematics.	No	Keep
10	I'm not the type to do well in math.	Reverse	Keep
11	Generally I have felt secure about attempting mathematics.	No	Keep
14	Math was my worst subject.	Reverse	Keep
15	I have a lot of self-confidence when it comes to math.	No	Keep
16	I am sure I could do advanced work in mathematics.	No	Remove
18	I think I could handle more difficult mathematics.	No	Keep
24	I'm no good at math.	Reverse	Keep
27	I can get good grades in mathematics.	No	Keep
35	For some reason even though I study, math seems unusually hard for me.	Reverse	Keep
Effectance Motivation in Mathematics Scale:			
1	I like math puzzles.	No	Remove
4	When a question is left unanswered in math class, I continue to think about it afterward.	No	Keep
6	Mathematics is enjoyable and stimulating to me.	No	Keep
9	I don't understand how some people can spend so much time on math and seem to enjoy it.	Reverse	Keep
13	I am challenged by math problems I can't understand immediately.	No	Keep
17	The challenge of math problems does not appeal to me.	Reverse	Keep
26	Math puzzles are boring.	Reverse	Keep
29	Once I start trying to work on a math puzzle, I find it hard to stop.	No	Keep
30	Figuring out mathematical problems does not appeal to me.	Reverse	Keep
31	When a math problem arises that I can't immediately solve, I stick with it until I have the solution.	No	Keep
33	I would rather have someone give me the solution to a difficult math problem than to have to work it out for myself.	Reverse	Keep
36	I do as little work in math as possible.	Reverse	Keep
Mathematics Anxiety Scale:			
3	I almost never get nervous during a math test.	Reverse	Keep
7	My mind goes blank and I am unable to think clearly when doing mathematics.	No	Keep
12	I usually have been at ease during math tests.	Reverse	Keep
19	Mathematics makes me feel uncomfortable, restless, irritable, or impatient.	No	Keep
20	I don't usually worry about being able to solve math problems.	Reverse	Keep
21	I usually have been at ease in math classes.	Reverse	Keep
22	I get a sinking feeling when I think of trying hard math problems.	No	Keep
23	A math test would scare me.	No	Keep
25	Mathematics usually makes me feel uncomfortable or nervous.	No	Keep
28	It wouldn't bother me at all to take more math courses.	Reverse	Keep
32	Mathematics makes me feel uneasy or confused.	No	Keep
34	Math doesn't scare me at all.	Reverse	Keep

References

Alexander, L., & Martray, C. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development, 22*(3), 143–150.

Alkhateeb, H. M. (2004). Internal consistency reliability and construct validity of an Arabic translation of the shortened form of the Fennema-Sherman mathematics attitudes scales. *Psychological Reports, 94*(2), 565–571.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology. General, 130*(2), 224–237.

Bai, H., Wang, L. S., Pan, W., & Frey, M. (2009). Measuring mathematics anxiety: psychometric analysis of a bidimensional affective scale. *Journal of Instructional Psychology, 36*(3), 185–193.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17*(5), 339–343.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 107*(5), 1860–1863.

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology, 25*(5), 441–448.

Bhargava, R. P., & Ishizuka, T. (1981). Selection of a subset of variables from the viewpoint of variation—an alternative to principal component analysis. *Proceedings of the Indian Statistics Institute Golden Jubilee International conference on statistics: applications and new directions*, (pp. 33–34). Calcutta: Indian Statistical Institute.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and pre-service elementary teachers' confidence to teach mathematics and science. *School Science and Mathematics, 106*(4), 173–179.

Chouinard, R., Vezeau, C., & Bouffard, T. (2008). Coeducational or single-sex school: does it make a difference on high school girls' academic motivation? *Educational Studies, 34*(2), 129–144.

Cooper, S. E., & Robinson, D. A. (1991). The relationship of mathematics self-efficacy beliefs to mathematics anxiety and performance. *Measurement and Evaluation in Counseling and Development, 24*(1), 4–11.

Dew, K. M. H., Galassi, J. P., & Galassi, M. D. (1983). Mathematics anxiety: some basic issues. *Journal of Counseling Psychology, 30*(3), 443–446.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). New York: Macmillan.

Fennema, E., & Sherman, J. A. (1976a). Fennema-Sherman Mathematics Attitude Scale: instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7*(5), 324–326.

Fennema, E., & Sherman, J. A. (1976b). Fennema-Sherman Mathematics Attitudes Scales. *JSAS: Catalogue of Selected Documents in Psychology, 6*(1), 31. (Ms. No. 1225).

Forgasz, H. J., Leder, G. C., & Gardner, P. L. (1999). The Fennema-Sherman Mathematics as a Male Domain scale reexamined. *Journal for Research in Mathematics Education, 30*(3), 342–348.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91.

Goldin, G. A. (2002). Affect, meta-affect, and mathematical belief structures. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: a hidden variable in mathematics education* (pp. 59–72). Dordrecht: Kluwer.

Goldring, R., Gray, L., & Bitterman, A. (2013*). Characteristics of public and private elementary and secondary school teachers in the United States: results from the 2011–12 Schools and Staffing Survey (NCES 2013–314).* U.S. Department of Education. Washington, DC: National Center for Education Statistics. http://nces.ed.gov/pubsearch. Accessed 7 Dec 2015.

Graven, M. (2004). Investigating mathematics teacher learning within an in-service community of practice: the centrality of confidence. *Educational Studies in Mathematics, 57*(2), 177–211.

Harper, N. W., & Daane, C. J. (1998). Causes and reductions of math anxiety in preservice elementary teachers. *Action in Teacher Education, 19*(4), 29–38.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58*(6), 6–11.

Hill, H., Blunk, L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cognition and Instruction, 26*(4), 430–511.

Hodges, T. E., & Jong, C. (2012). Exploring changes in preservice teachers' conceptions within the context of mathematics experiences. In Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds), *Proceedings of the 34rd*

Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, (pp. 709–712). Kalamazoo: Western Michigan University.

Jackson, D. L., Gillaspy, J. J., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods, 14*(1), 6–23.

Jackson, E. (2015). Student primary teachers' perceptions of mathematics. *Philosophy of Mathematics Education Journal, 29.* http://people.exeter.ac.uk/PErnest/pome29/index.html. Accessed 7 Dec 2015.

Jong, C., Hodges, T. E., & Welder, R. M. (2012). *Conceptions of mathematics in related contexts: measuring elementary teachers' development over time. Paper presented at the American Educational Research Association conference.* Canada: Vancouver.

Karp, K. S. (1991). Elementary school teachers' attitudes toward mathematics: the impact on students' autonomous learning skills. *School Science and Mathematics, 91*(6), 265–270.

Leder, G. C., & Grootenboer, P. J. (2005). Affect in mathematics education. *Mathematics Education Research Journal, 17*(2), 1–8.

Lee, J. (2005). Correlations between kindergarten teachers' attitudes toward mathematics and teaching practice. *Journal of Early Childhood Teacher Education, 25*(2), 173–184.

Lim, S. Y., & Chapman, E. (2013). An investigation of the Fennema-Sherman Mathematics Anxiety Subscale. *Measurement and Evaluation in Counseling and Development, 46*(1), 26–37.

Lipnevich, A. A., MacCann, C., Krumm, S., Burrus, J., & Roberts, R. D. (2011). Mathematics attitudes and mathematics outcomes of U.S. and Belarusian middle school students. *Journal of Educational Psychology, 103*(1), 105–118.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520–540.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: a meta-analysis. *Journal for Research in Mathematics Education, 28*(1), 26–47.

McAnallen, R. R. (2010). *Examining mathematics anxiety in elementary classroom teachers* (unpublished doctoral dissertation). http://www.gifted.uconn.edu/siegle/Dissertations/Rachel%20McAnallen.pdf. Accessed 7 Dec 2015.

McLeod, D. B. (1992). Research on affect in mathematics education: a reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: Macmillan.

Melanchon, J. G., Thompson, B., & Becnel, S. (1994). Measurement integrity of scores from the Fennema-Sherman Mathematics Attitudes Scales: the attitudes of public school teachers. *Educational and Psychological Measurement, 54*(1), 187–192.

Muis, K. R., & Foy, M. J. (2010). The effects of teachers' beliefs on elementary students' beliefs, motivation, and achievement in mathematics. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal epistemology in the classroom: theory, research, and implications for practice* (pp. 435–469). Cambridge: Cambridge University Press.

Mulhern, F., & Rae, G. (1998). Development of a shortened form of the Fennema-Sherman Mathematics Attitudes Scales. *Educational and Psychological Measurement, 58*(2), 295–306.

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J. (2008). TIMSS 2007 international mathematics report: findings from IEA's trends in international mathematics and science study at the fourth and eighth grade. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College. http://pirls.bc.edu/timss2007/mathreport.html. Accessed 7 Dec 2015

Muthén, L. K., & Muthén, L. (1998–2013). *Mplus [computer software].* Los Angeles: Muthén & Muthén.

Neale, D. C. (1969). The role of attitudes in learning mathematics. *Arithmetic Teacher, 16*, 631–640.

Norton, S. J., & Rennie, L. J. (1998). Students' attitudes towards mathematics in single-sex and coeducational schools. *Mathematics Education Research Journal, 10*(1), 16–36.

Pajares, M. F. (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.

Pearn, C., Brew, C., Leder, G., & Bishop, A. (1996). Attitudes towards mathematics: what about NESB students? In P. C. Clarkson (Ed.), *Technology in mathematics education: Proceedings o f Nineteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 445–452). Melbourne, Australia: Mathematics Education Research Group of Australia Inc.

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Charlotte: Information Age Publishing.

Quinn, R. J. (1997). Effects of mathematics methods courses on the mathematical attitudes and content knowledge of preservice teachers. *Journal of Educational Research, 92*(2), 108–113.

Reyes, L. H. (1984). Affective variables and mathematics education. *The Elementary School Journal, 84*(5), 558–581.

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: psychometric data. *Journal of Counseling Psychology, 19*(6), 551–554.

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

Rowe, K. J. (1993). What are the benefits of single-sex maths classes? In I. Livingstone & J. Izard (Eds.), *Best of SET mathematics*. Camberwell: Australian Council for Educational Research.

Sachs, J., & Leung, S. O. (2007). Shortened versions of Fennema-Sherman Mathematic Attitude Scales employing trace information. *Psychologia: An International Journal of Psychology in the Orient, 50*(3), 224–235.

Sherman, H. J., & Christian, M. (1999). Mathematics attitudes and global self-concept: an investigation of the relationship. *College Student Journal, 33*(1), 95–101.

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

Swars, S., Daane, C., & Giesen, J. (2006). Mathematics anxiety and mathematics teacher efficacy: what is the relationship in elementary preservice teachers? *School Science and Mathematics, 106*(7), 306–315.

Tapia, M., & Marsh, G. E., II. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly, 8*(2), 16–21.

Tatto, M. T. (Ed.). (2013). *The Teacher Education and Development Study in Mathematics (TEDS-M): policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Technical report.* Amsterdam: International Association for the Evaluation of Education Achievement (IEA).

The World Bank. (2015). *Primary education, teachers (% female)*. http://data.worldbank.org/indicator/SE. PRM.TCHR.FE.ZS. Accessed 7 Dec 2015.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: a synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York: MacMillan.

Tsao, Y. (2014). Attitudes and beliefs toward mathematics for elementary pre-service teachers. *US-China Education Review B, 4*(9), 616–626.

Vezeau, C., Chouinard, R., Bouffard, T., & Couture, N. (1998). Adaptation et validation des échelles de Fennema-Sherman sur les attitudes en mathématiques chez des garçons et des filles du secondaire. *Revue canadienne des sciences du comportement, 30*(1), 137–140.

Welder, R. M., & Jong, C. (2012). Examining connections between mathematical knowledge for teaching and conceptions about mathematics teaching and learning. In L. R.Van Zoest, J.-J. Lo, & J. L. Kratky. (Eds), *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (pp. 773–776). Kalamazoo: Western Michigan University.

Welder, R. M., Hodges, T. E., Jong, C. (2011). Measuring changes in teachers' beliefs, attitudes, and dispositions related to experiences in mathematics. In L. R. Wiest & T. Lamberg. (Eds.), *Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (pp. 2118–2125). Reno: University of Nevada, Reno.

Wilkins, J. L. M. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139–164.

Wilkins, J. L. M. (2010). Elementary school teachers' attitudes toward different subjects. *The Teacher Educator, 45*(1), 23–36.