Received: 3 May 2022 | Revised: 3 November 2022 | Accepted: 11 November 2022

DOI: 10.1111/ssm.12558

RESEARCH ARTICLE

WILEY | SSMA

Validating the use of student-level instruments to examine preservice teachers' mathematical problem solving

Timothy D. Folger[1] | **Maria Stewart**[2] | **Jonathan Bostic**[3] | **Toni A. May**[4]

[1]School of Educational Foundations, Leadership and Policy, Bowling Green State University, Bowling Green, Ohio, USA

[2]Department of Learning, Teaching and Curriculum, University of Missouri, Columbia, Missouri, USA

[3]School of Teaching and Learning, Bowling Green State University, Bowling Green, Ohio, USA

[4]School of Educational, Drexel University, Philadelphia, Pennsylvania, USA

Correspondence
Timothy D. Folger, School of Educational Foundations, Leadership and Policy, Bowling Green State University, Bowling Green, Ohio, USA
Email: tdfolge@bgsu.edu

Funding information
National Science Foundation, Grant/Award Numbers: 1720646, 1720661, 1920619, 1920621

Abstract

Problem solving is a central focus of mathematics teaching and learning. If teachers are expected to support students' problem-solving development, then it reasons that teachers should also be able to solve problems aligned to grade level content standards. The purpose of this validation study is twofold: (1) to present evidence supporting the use of the Problem Solving Measures Grades 3–5 with preservice teachers (PSTs), and (2) to examine PSTs' abilities to solve problems aligned to grades 3–5 academic content standards. This study used Rasch measurement techniques to support psychometric analysis of the Problem Solving Measures when used with PSTs. Results indicate the Problem Solving Measures are appropriate for use with PSTs, and PSTs' performance on the Problem Solving Measures differed between first-year PSTs and end-of-program PSTs. Implications include program evaluation and the potential benefits of using K-12 student-level assessments as measures of PSTs' content knowledge.

KEYWORDS

learning processes, math/math education, problem solving, teachers and teaching, teacher education, teacher knowledge

1 | INTRODUCTION

Effective mathematics teaching and learning engages students in tasks that promote reasoning and problem solving (Association of Mathematics Teacher Educators [AMTE], 2017; National Council of Teachers of Mathematics [NCTM], 2014). Students' classroom experiences with mathematical problem solving; however, depends on teachers' knowledge, beliefs, and attitudes of mathematics (Wilkins, 2008). Teachers' knowledge of mathematics impacts decisions about the mathematical tasks, instructional scaffolds, and mathematical discourse occurring during a classroom lesson (Curcio & Artzt, 2003; Schoenfeld, 2011). Such decision-making is imperative in supporting each and every student as a major facet of students' understanding and in turn, teaching, is mathematical problem solving (NCTM, 2000). Problem solving is characterized as the act of navigating a challenging situation and finding a solution to a problem (NCTM, 2000). We define mathematical problem solving as, "the process of interpreting a situation mathematically, which usually involves several cycles of expression, testing, and revising mathematical interpretations" (Lesh & Zawojewski, 2007, p. 782). One issue teachers face is how to engage their students in mathematical problem solving during classroom instruction. Consequently, teacher preparation programs should consider whether preservice teachers possess the knowledge and skills needed to lead mathematics instruction through a problem-solving approach.

The purpose of this study is twofold: (a) To investigate whether prior mathematical problem-solving measures,

which were validated for use with students in grades 3–5, are appropriate for use with preservice teachers, and (b) To examine preservice teachers' (PSTs) ability to solve mathematical word problems developed for students they might teach. We draw upon a prior claim, "If PSTs are expected to design and lead instruction focused on [problem solving], then it follows they should be able to solve problems related to the content standards" (Nielsen & Bostic, 2020, p. 35). As such, this study uses the Problem Solving Measure (PSM) series, initially designed for grades 3, 4, and 5 students, to measure PSTs' problem-solving performance. As the AMTE (2017) standards suggest, novice teachers should be able to solve the problems their students solve. There were two research questions for this study.

- (RQ1) What are the psychometric properties for the PSM3, PSM4, and PSM5 when used with preservice teachers?
- (RQ2) Are there significant differences between first-year and end-of-program PSTs' problem-solving performance using the PSMs?

2 | RELATED LITERATURE

2.1 | Problem solving framework

Problem solving is an important feature of mathematics teaching and learning (NCTM, 2000; 2014). Mathematics word problems are frequently a central aspect of mathematics instruction that intends to promote problem solving (Bostic et al., 2016; Palm, 2006, 2008; Reed, 1998; Verschaffel et al., 2000). Problem solving inherently requires a task that is a problem. A *problem* (a) lacks an apparent solution strategy and (b) contains multiple viable solution strategies (NCTM, 2014; Schoenfeld, 2011). Problems differ from exercises. An *exercise* is a task meant to promote proficiency with a known procedure (Kilpatrick et al., 2001; Mayer & Wittrock, 2006). The problem-solving framework set forth in this study is the intersection of the presented definition of a problem and Verschaffel et al. (1999) categorization of word problems as (a) open, (b) developmentally complex, and (c) realistic. Open problems can be solved using more than one developmentally appropriate strategy. Problems are developmentally complex when a solution strategy is not apparent (Schoenfeld, 2011). Realistic tasks contain a believable situational context drawn from real-life experiences (Verschaffel et al., 1999). Therefore, we define word-problems as a mathematical problem situated within a believable real-life context.

Problem solving permeates the Common Core State Standards for Mathematical Content (SMCs); it is also highlighted in its own Standard for Mathematical Practice (e.g., Standard for Mathematical Practice 1; National Governors Association & Council of Chief State School Officers, 2010). In the Common Core State Standards for Mathematics, there is at least one content standard in every grade level from K-5 that has some mention of, "solve problems involving" (National Governors Association & Council of Chief State School Officers, 2010). Table 1 provides some examples. Note that the standards specifically say to solve problems, and not exercises. Furthermore, many standards refer to real-world problems, which aligns with our definition for word problems. This reference to solving problems in the content standards paired with the Standards for Mathematical Practice clearly emphasizes the importance of problem solving as a critical mathematical experience for students.

TABLE 1 Examples of standards that emphasize problem solving

Grade level	Standards
3	3.MD.D.8: Solve real world and mathematical problems involving perimeters of polygons, including finding the perimeter given the side lengths, finding an unknown side length, and exhibiting rectangles with the same perimeter and different areas or with the same area and different perimeters
4	4.NF.B.3.D: Solve word problems involving addition and subtraction of fractions referring to the same whole and having like denominators, e.g., by using visual fraction models and equations to represent the problem
5	5.NF.B.7.C: Solve real world problems involving division of unit fractions by nonzero whole numbers and division of whole numbers by unit fractions, e.g., by using visual fraction models and equations to represent the problem

2.2 | Effective teaching and learning of mathematics—AMTE/NCTM standards

Teachers are expected to teach students mathematics content and promote mathematical behaviors and habits; therefore, they should have competency with these bodies of knowledge. The AMTE (2017) Standards for Preparing Teachers of Mathematics were created to provide teacher-preparation programs with a set of standards for preparing high-quality teachers of mathematics who can support students using best teaching practices. One section of the standards focuses on "candidate knowledge, skills, and dispositions," which preservice teachers need to be

successful teachers upon graduation (AMTE, 2017, p. 5). Standard C.1 states, "well-prepared beginning teachers of mathematics possess robust knowledge of mathematical and statistical concepts that underlie what they encounter in teaching" (AMTE, 2017, p. 8). This standard emphasizes the importance that PSTs and teachers alike must possess the skills and knowledge necessary to do the mathematics they will eventually engage in with their students. Since the Common Core State Standards for Mathematics (CCSSI, 2010) focus on problem solving, PSTs and novice teachers should be able to demonstrate proficiency as problem solvers engaging with mathematics content they are expected to teach. AMTE (2017) standards also support the idea that PSTs should be good problem solvers stating in standard C.1.2, "[PSTs] can apply their mathematical knowledge to real-world situations by using mathematical modeling to solve problems appropriate for the grade levels and the students they will teach" (p. 9). This standard supports the idea that PSTs and novice teachers should be able to solve the problems they will be asking their students to solve. Teachers' ability to solve such problems could be measured by mathematical problem-solving assessments designed for elementary students, such as the Problem Solving Measures for grades 3–5.

2.3 | Teachers' knowledge of mathematics

Mathematical content knowledge (MCK) refers to knowledge of mathematics and mathematical structures (Ball et al., 2008; Shulman, 1986). MCK is one element of teachers' knowledge. Shulman (1986) explains, "content knowledge requires going beyond knowledge of the facts or concepts of a domain. It requires understanding the structure of the subject matter" (p. 9). Therefore, MCK extends beyond mathematical performance or calculation, and includes a conceptual understanding of the subject matter. For this study, MCK includes common content knowledge, which is mathematical knowledge used outside of the teaching setting, and specialized content knowledge for mathematics, which is the mathematical knowledge needed for teaching (Ball et al., 2008). Teachers' knowledge of mathematics has direct implications for instructional practices. Gains in teachers' content knowledge for teaching correlate with improvements in instructional quality and classroom climate (Copur-Gencturk, 2015). Furthermore, Dunekacke et al. (2015) state, "Prospective... teachers with more mathematics content knowledge perceive situations... that are related to mathematics on average more precisely than teachers with less knowledge" (pp. 280–281). That is, there is a need for teachers to understand the mathematical structures at work to reify mathematical concepts for students.

Extensive research has been conducted regarding teachers' knowledge of mathematics (e.g., Campbell et al., 2014; Hill et al., 2004; White et al., 2013; Wilkins, 2008). White et al. (2013) reported gains in inservice teachers' mathematical knowledge for teaching following participation in a Math Teachers' Circle, which regularly engaged participants in mathematical problem-solving activities. Their study demonstrated how problem-solving experiences can be positively related to improvement in mathematical content knowledge. Such gains affirm a relationship between mathematical problem solving and mathematical understanding. Therefore, understanding PSTs' problem-solving performance, which is certainly informed by their content knowledge, provides a window through which mathematics teacher educators may consider PSTs readiness to facilitate mathematics problem-solving activities with their future students.

Content knowledge and problem solving are inextricably connected (Lambdin, 2003); "The connection between solving problems and deepening reasoning is symbiotic" (p. 6). Knowledge is a necessary component for problem-solving success (Schoenfeld, 2011). Meanwhile, teaching through problem solving promotes robust mathematical understanding (Bostic et al., 2016; Lambdin, 2003; Schroeder & Lester, 1989). In other words, while a proficient level of content knowledge is needed to successfully solve problems, effectively scaffolded problem-solving instruction also deepens students' content knowledge (Bostic et al., 2016; Curcio & Artzt, 2003; Lambdin, 2003; Schoenfeld, 2011). Mathematical problem solving and mathematical understanding possess a reciprocal relationship; understanding supports problem solving, and problem solving then deepens understanding.

Consequently, PSTs should (a) possess the mathematical content knowledge needed to support students'' mathematical problem solving and (b) be able to apply their mathematical content knowledge to solve problems related to the content they may be required to teach. The research within teacher education literature on PSTs'' mathematical problem solving is somewhat fractured in this area. A prior study conducted by Nielsen and Bostic (2020) examined first- and fourth-year secondary mathematics PSTs'' problem-solving performance with grades 6–-8 content. Results from that study indicated PSTs struggled to solve mathematics word problems aligned to the content they may be required to teach. A key implication from that study was the importance of interdepartmental conversations about secondary PSTs'' mathematics content experiences. Specifically, the need to include more problem-solving opportunities for PSTs during their undergraduate experience (Nielsen & Bostic, 2020). That study started to fill a

gap in the literature, which may be partially credited to a lack of instruments in mathematics education designed and validated to make interpretations about test-takers'' problem-solving ability (Bostic & Sondergeld, 2015).

2.4 | Validity, validation, and the problem solving measures

Validity is defined as the degree to which the interpretation of test scores for an intended use is supported by research and theory (American Education Research Association [AERA] et al., 2014). Validity is a unitary concept and is an attribute of the interpretation(s) and use(s) of test scores (AERA et al., 2014; Kane, 2013). *The Standards for Educational and Psychological Testing [The Standards]* outline five sources of validity evidence: test content, response processes, internal structure, relations to other variables, and consequences of testing (AERA et al., 2014). Table 2 displays a brief description of each source of validity evidence. Not all five validity sources are required to establish a degree of validity (AERA et al., 2014; Kane, 2016). Each type of evidence supports differing claims inherent to the interpretation and use of test scores. For instance, evidence based on relationships to other variables may support claims that the test taker's performance is related to some external criterion, such as future academic achievement. Decisions about the type and quantity of evidence to present depends on the complexity of how test scores are interpreted and used. Complex interpretations and uses of test scores require a more robust validity argument (AERA et al., 2014; Kane, 2013). For example, a mathematics achievement test may have sufficient validity evidence presented to draw inferences regarding students' knowledge of grade-level content standards, but insufficient evidence to draw inferences regarding the quality of instruction provided by the test-taker's teacher. Research has indicated the mathematics education community has not necessarily adhered to an argument-based approach to validation (Bostic et al., 2021; Carney et al., 2022).

The PSMs, developed for use in grades 3–8, are an assessment of problem-solving performance in relation to respective grade-level Common Core State Standards. There is burgeoning evidence of validity supporting the interpretation of PSM results as students' problem-solving performance aligned to grade-level content standards (e.g., Bostic, 2018; Bostic et al., 2017; Bostic & Sondergeld, 2015). For instance, validity evidence based on test content were collected for each PSM through expert panel reviews of items (e.g., Bostic, 2018; Bostic et al., 2017; Bostic & Sondergeld, 2015). Evidence based on test content supports the claim that PSMs assess students' ability to solve problems aligned to the Common Core State Standards for Mathematics (CCSSM). Furthermore, validity evidence based on response processes has been reported in previous research (e.g., Bostic, 2018; Bostic & Sondergeld, 2015; Bostic et al., 2017). This prior research reports on think-alouds conducted with grade-level students, concluding that (a) students are able to read and interpret the situational context of the items, and (b) evidence of students' productive struggle in solving items aligns with the problem-solving framework. A comprehensive review of PSM development and validation for use with grades 3–8 students is beyond the scope of this manuscript; readers interested in that information may refer to Bostic et al. (2022), Bostic (2018), Bostic et al. (2017), and Bostic and Sondergeld (2015). The following section presents an interpretation and use statement (IUS) using guidelines recommended by Carney et al. (2022), which details the intended

TABLE 2 Descriptions of the five sources of validity evidence

Source of evidence	Description
Test content	"Test content refers to the themes, wording, and format of the items, tasks, or questions on a test" (AERA et al., 2014, p. 14)
Response processes	"Theoretical and empirical analyses of the response processes of test takers can provide evidence concerning the fit between the construct and the detailed nature of the performance or response actually engaged in by test takers." (AERA et al., 2014, p. 15)
Internal structure	"Analyses of the internal structure of a test can indicate the degree to which the relationships among the items and test components conform to the construct on which the proposed test score interpretations are based" (AERA et al., 2014, p.16)
Relations to other variables	Relations to other variables may provide evidence, for example, that indicates how "...test scores [may or may not be] influenced by ancillary variables such as [individual or group characteristic]" (AERA et al., 2014, p.12)
Consequences of testing	"...decisions about test use are appropriately informed by validity evidence about intended test score interpretations for a given use, by evidence evaluating additional claims about consequences of test use that do not follow directly from test score interpretations, and by value judgments about unintended positive and negative consequences of test use." (AERA et al., 2014, p. 21)

interpretation of PSM scores when used with PSTs. An IUS is organized into three overarching categories: (a) construct articulation, (b) operationalization and administration, and (c) scores and usage.

2.4.1 | Construct articulation

In articulating the construct, the PSMs are an assessment of problem-solving performance in relation to respective grade-level Common Core State Standards. We acknowledge that problem-solving performance is influenced by content knowledge and developmental readiness. That is, a task classified as a problem for most individuals may be a routine exercise for an individual with an advanced knowledge of the mathematics content. The biggest concern when using the PSMs with PSTs was if the items were developmentally complex for the PSTs. A prior study (Nielsen & Bostic, 2020) found that sixth, seventh, and eighth grade PSM items used with secondary (grades 7–12) PSTs were developmentally complex. The PSTs in that study had an average score of 77%, 67%, and 54% respectively, suggesting that the PSTs found the PSM items challenging. Thus, although PSM items should be less cognitively demanding for PSTs than grades 3–5 students, the PSM items are still challenging problems such that solution strategies are not readily apparent to a typical student or preservice teacher. In this study, we argue the construct is important to measure with PSTs because they may eventually be responsible for providing instruction centered around mathematical problem-solving.

2.4.2 | Operationalization and administration

Each grade-level test consists of 15–19 constructed-response items. Items are developed using the described problem-solving framework. That is, word problems appearing on the PSMs are classified as problems, not exercises, for the typical grade-level student. The typical student is the unit of analysis because it is feasible for an item to be a problem for one particular student, but that same item may be an exercise for a different student depending on the developmental ability of each student. PSM items are classified as open based on the existence of multiple solution strategies. With exception to the PSM3, which is the earliest grade-level test in the series, each PSM contains at least three common linking items shared with the previous grade-level test. For instance, the PSM4 contains three items that were developed for the PSM3. All remaining items are aligned to the respective grade-level content standards. This study identifies PSTs as the target population with whom the PSMs are administered. PSTs should be provided ample time to complete the PSM, but are not permitted to use a calculator during test administration.

2.4.3 | Scores and usage

This study follows past practice (see Bostic et al., 2017; Bostic & Sondergeld, 2015) for scoring the PSMs. Participants' responses on the PSM items are scored dichotomously as correct or incorrect, and this study evaluates PSTs' performance through Rasch (1960) modeling. Rasch person measures (i.e., logit scores) can be interpreted as the relative probability of PSTs producing a correct response to CCSSM-aligned word problems based on item difficulty (Bond & Fox, 2015). PSTs with greater Rasch person measures have a greater likelihood of success when solving PSM items. Raw scores can also be used to describe PST performance (Nielsen & Bostic, 2020). It should be cautioned that the use of raw scores does not reflect the hierarchy of item difficulty inherent to each PSM, but raw scores can still be interpreted to draw inferences of PST problem-solving performance. PSM scores should only be used for low-stakes purposes. In this study, we propose using PSTs' PSM scores formatively to inform the types of mathematics tasks and activities PSTs are exposed to in their teacher-preparation courses. Subsequently, we propose using PSM scores as an element to evaluate teacher-preparation programs. Mathematics teacher educators may use scores to evaluate the degree to which their teacher-preparation program is fostering PSTs' ability to solve mathematical problems aligned to content they may be required to teach. In interpreting PSTs' PSM scores, it should be cautioned that content knowledge and problem-solving are necessary, but not sufficient conditions for being an effective teacher. To be clear, PSM scores are not predictive of PSTs' overall teaching ability or future teaching quality.

This validation study presents evidence based on internal structure and evidence based on relations to other variables supporting the use of the PSMs with PSTs. Evidence based on internal structure were gathered through the psychometric examination of the PSMs (i.e., RQ1). Evidence based on relations to other variables were gathered by examining the hypothesized differences between first-year and end-of-program PSTs' problem-solving performance (i.e., RQ2).

3 | METHODS

3.1 | Research design and context

This validation study took place at a large public university in the Midwest. The public university enrolls

TABLE 3 PST participation based on PSM and program status

	Participation prior to data screening		Participation following data screening	
	First-year IEC majors	End-of-program IEC majors	First-year IEC majors	End-of-program IEC majors
PSM3	104	88	93	81
PSM4	101	93	101	93
PSM5	100	95	98	94
Total	305	276	292	268

relatively large cohorts of undergraduate students in various teacher preparation programs. A cohort is defined as a group of students who enrolled in their teacher preparation program at the same point in time. Participants include preservice teachers (PSTs) enrolled in the inclusive early childhood (IEC) teacher preparation program. The IEC program leads to state licensure to teach students in prekindergarten through grade five.

3.1.1 | Participants and procedures

Four different cohorts of students are represented in this study. Participants included two cohorts of first-year PSTs and two cohorts of end-of-program IEC PSTs. Participants were enrolled in a four-year bachelor of science degree in IEC education. A total of 581 PSTs completed one PSM. Personal demographic data were not collected during PSM administration; however, more than 90% of PSTs in the IEC teacher preparation program are classified as white females. A unique characteristic of these participants is that, in large, their K-12 educational experiences as students should have aligned with the CCSSM standards, which have a general focus on developing mathematical problem solving abilities (National Governors Association & Council of Chief State School Officers, 2010). More specifically, participants would have been enrolled in grades 2–5 when the Common Core State Standards were adopted in 2010. To be clear, the large majority of participants would have attended K-12 schools in a state that had adopted CCSS in 2010, but this may not be true for all participants. The PSM3, PSM4, and PSM5 were used to measure PSTs' ability to solve word problems aligned to content they may be required to teach. Each student completed one PSM. Table 3 displays the number of PSTs who completed their assigned PSM, disaggregated by PSM and PSTs' status in their teacher preparation program. Preservice teachers were randomly administered a PSM while enrolled in one of two courses exploring the teaching and learning of elementary mathematics content. First-year PSTs were enrolled in an introductory mathematics education course. End-of-program PSTs were enrolled in a mathematics methods course for IEC PSTs. Data were collected between September of 2020 and September of 2021, during the COVID-19 pandemic. Students completed their assigned PSM online and outside of class time using Google Forms.

Data were screened prior to and concurrent with Rasch analysis and outliers were removed. Person point-biserial indices, and infit and outfit mean square (MNSQ) statistics were used to help identify eccentric individuals who may or may not have taken the assessment seriously (Boone & Noltemeyer, 2017). For example, submissions that contained responses such as "IDK" or "I don't know" rather than attempting to respond to each item were identified and removed. As a result, 21 participants were removed from the study for a grand total of 560 participants. Table 3 also displays the number of PSTs' that were included following data screening.

4 | DATA ANALYSIS

4.1 | Validity evidence: Internal structure

Validity evidence based on internal structure may be produced through analysis of an instrument's dimensionality (Rios & Wells, 2014). Evidence of unidimensionality supports the claim that the PSMs measure a single construct (i.e., problem-solving performance in relation to respective grade-level Common Core State Standards). Data were examined using Rasch (1960) measurement for dichotomous responses. Rasch measurement is one method to establish validity evidence based on the internal structure of the instrument (Bostic et al., 2017; Bostic & Sondergeld, 2015; Smith Jr, 2002). This approach constructs a linear statistical model from observed counts and categorical responses (Wright & Stone, 1999). Unidimensionality is a requirement for any measurement model. However, it is important to distinguish between theoretical unidimensionality and practical or functional unidimensionality. Functional unidimensionality acknowledges that several constructs may work together in measuring a latent trait (Smith, 1996). For instance, factor analyses of arithmetic tests often identify four factors: addition, subtraction, multiplication, and division. Yet, those four closely related factors are framed as measuring a single construct: arithmetic (Smith, 1996). Unidimensionality cannot be assessed dichotomously as unidimensional or not, and there is no best approach to examine unidimensionality (Smith Jr, 2002). Common

across scholarship discussing Rasch measurement is the idea that "the unidimensionality requirement is satisfied when the data fit the model" (Smith, 1996, p. 26). Such a definition posits an emphasis on fit statistics for items and persons. We utilized several methods to holistically examine the unidimensionality of the PSMs 3–5: item fit, item point-biserial correlations, as well as item separation and reliability.

Mean-square (MNSQ) infit and outfit statistics were examined at both the item- and person-level (Boone & Noltemeyer, 2017; Smith, 1996). Item fit describes the degree to which responses for each particular item aligns with the Rasch model expectation (Wright & Stone, 1999). Person fit summarizes the degree to which that respondent's pattern of performance aligns with how respondents typically perform (Wright & Stone, 1999). Infit examines patterns of performance when item difficulty and person ability are similar. Outfit examines patterns of performance when item difficulty and person ability are far apart (Linacre, 2002). Fit statistics between 0.5 and 1.5 are acceptable for low stakes assessments (Wright & Linacre, 1994).

Corrected point-biserial correlations were analyzed to measure how items function in relation to one another. Correlation indices range from −1 to 1. Items producing a negative point-biserial are a concern and should be considered for removal from the analysis because such items fail to represent the latent trait being measured by the instrument (Wright, 1992). Logically, items with a negative point-biserial correlation pull in a direction opposite of the other items.

Rasch item separation and reliability were examined to inform on the dimensionality of the PSMs. Item separation indicates a hierarchy of item difficulty. More specifically, item separation is used to identify the number of statistically distinct groups (i.e., strata) regarding item difficulty. For example, a separation statistic of 2.0 equates to a strata of 3, or 3 statistically distinct groups. These three groups could be interpreted as easy items, moderate items, and challenging items. Item reliability ranges from 0 to 1.00. Rasch item reliability is similar to traditional measures of reliability (i.e., Cronbach's alpha), but indicates consistency in item difficulty rather than person performance. Item reliability and separation indices are respectively classified as excellent at 0.90 and 3.0, good at 0.80 and 2.00, and acceptable at 0.70 and 1.50 (Duncan et al., 2003). Item invariance was considered by comparing subsample performance on common linking items. Common items across the PSMs link data sets together. The invariance principle posits that relative item difficulty should remain consistent across subsamples of test-takers (Bond & Fox, 2015). In this case, PSTs were randomly assigned a PSM to complete; therefore, item difficulty should remain invariant across PST subsamples and the grade level test(s) containing said item. Data were analyzed using Winsteps version 3.74 (Linacre, 2012) to examine the internal structure of the PSMs 3–5.

4.1.1 | Validity evidence: Relationships to other variables

PSTs' performance based on program year was examined to collect validity evidence based on relations to other variables. We hypothesized that end-of-program PSTs might score statistically significantly greater than first-year PSTs on the PSMs. This hypothesis was informed by two ideas. First, prior research with the PSMs for grade 6–8 indicated such a relationship. Second, end-of-program PSTs have likely been exposed to more opportunities to engage in mathematical problem-solving activities than first-year PSTs. Evidence of a statistically significant difference in problem-solving performance between first-year and end-of-program PSTs supports the claim that test-takers' performance aligns with our expectations - that performance is related to program year.

Rasch analysis person measures (i.e., logit scores) and PSTs' raw scores were used to analyze PSTs' performance on the PSM3, PSM4, and PSM5. Rasch person and item measures are reported in logit units which describe a relative amount of the latent trait. A benefit of Rasch analysis is that logit units are expressed on a linear scale and can be used in subsequent statistical analyses (Bond & Fox, 2015; Boone & Noltemeyer, 2017). Person measures were compared to the mean item difficulty on each PSM to assess the overall challenge students faced. To establish validity evidence based on relationship to other variables, an ANOVA was conducted for each PSM using PSTs' logit scores to examine mean differences in scores based on program year. PSTs' logit scores were used in subsequent analyses over raw scores because logit scores are reflective of item difficulty and logit scores are linear. For example, two individuals that both correctly respond to 10 out of 15 PSM items appear to contain an equal amount of the latent trait when analyzing raw scores, but the logit score will be greater for the person who correctly answered more difficult items. Analysis of variance (ANOVA) for PSTs' logit scores were conducted using SPSS version 27.

5 | RESULTS

5.1 | Validity evidence: Internal structure

Our first research question was: What are the psychometric properties for the PSM3, PSM4, and PSM5 when used

with preservice teachers? Rasch infit and outfit MNSQ indices were appropriate for all PSM items. Table 5 displays the range of infit and outfit statistics for each PSM. For low-stakes assessments, MNSQ indices between 0.5 and 1.5 are deemed acceptable (Wright & Linacre, 1994). Values less than 0.5 are classified as overfitting. Overfitting items are not productive in contributing unique information about test takers, but they do not raise concerns about the unidimensionality of the instrument. MNSQ indices greater than 2.0 are classified as misfitting. Such items do raise concerns regarding the dimensionality of the instrument. As displayed in Table 4, no PSM items reported MNSQ infit or outfit values greater than 1.46. One PSM item had a MNSQ outfit value of 0.37 and was flagged as an overfitting item.

No items on any PSM produced a negative point-biserial statistic suggesting all items appear to be working together in measuring a single latent variable. Corrected point-biserial statistics ranged from 0.18 to 0.57 across all PSM items used in this study. Varma (2006) classifies item point-biserial values greater than 0.25 as good and values greater than 0.15 as acceptable; indicating some PSM items fall within an acceptable to good range. Thus, providing evidence that PSM items function well together in measuring PSTs' ability to solve mathematical problems aligned to the Common Core State Standards. Table 5 displays the item separation and item reliability for each of the PSMs used in this study. Separation and reliability indices were classified as excellent for each PSM (Duncan et al., 2003). Regarding item invariance, item measures for five of six linking items were found to be invariant. The difference of respective item measures (i.e., logit scores) between the subsamples of PSTs were within one standard error of the items, indicating PSM item difficulty is invariant across subsequent grade-level tests and the sample of students assigned to each PSM. However, there was one linking item between the PSM3 and PSM4 for which the item-measure difference exceeded one standard error of the items but fell within two standard errors of the items. More specifically, PSTs completing the PSM4 found the item easier than PSTs completing the PSM3. Taken collectively, findings from psychometric item analyses suggest the PSM3, PSM4, and PSM5 function reasonably well as unidimensional measures of preservice teachers' problem-solving performance.

TABLE 4 PSM infit and outfit statistics

	MNSQ infit statistics		MNSQ outfit statistics	
	Minimum	**Maximum**	**Minimum**	**Maximum**
PSM3	0.81	1.27	0.37	1.26
PSM4	0.85	1.15	0.73	1.34
PSM5	0.84	1.16	0.77	1.31

TABLE 5 Item separation and reliability

	PSM3	PSM4	PSM5
Item separation	4.48	3.92	4.63
Item reliability	0.95	0.94	0.96

5.2 | Validity evidence: Relations to other variables

Preliminary analysis of PSTs' problem-solving performance indicates that PSTs were most successful solving problems aligned to grade three content standards. Participants' mean raw score was 10.28 out of 14 and their average logit score was 1.51 on the PSM3. Additionally,

TABLE 6 Descriptive statistics

		PSTs' raw scores				PST's logit scores			
		Mean	***SD***	**Max**	**Min**	**Mean**	***SD***	**Max**	**Min**
PSM3	Total ($n = 174$)	10.28	2.35	14	4	1.51	1.31	4.47	−2.13
	First-year ($n = 93$)	9.78	2.27	14	4	1.24	1.22	4.47	−2.13
	End-of-program ($n = 81$)	10.77	2.33	14	4	1.83	1.36	4.47	−0.99
PSM4	Total ($n = 194$)	8.74	2.57	14	3	0.57	1.10	3.95	−2.95
	First-year ($n = 101$)	8.49	2.77	14	3	0.51	1.13	3.95	−2.95
	End-of-program ($n = 93$)	9.00	2.32	14	4	0.72	0.95	3.95	−1.13
PSM5	Total ($n = 192$)	9.54	2.78	15	3	0.84	1.14	4.37	−2.06
	First-year ($n = 98$)	9.02	2.87	14	3	0.62	1.07	3.07	−2.06
	End-of-program ($n = 94$)	10.04	2.70	15	3	1.07	1.17	4.37	−2.06

end-of-program PSTs scored higher than first-year PSTs on the PSM3, PSM4, and PSM5. Table 6 displays descriptive statistics for each PSM.

Our second research question was: Are there significant differences between first-year and end-of-program PSTs' problem-solving performance using the PSMs? ANOVA results indicate a statistically significant difference in PSTs' problem-solving performance based on program year on the PSM3, $F(1, 172) = 9.184, p = 0.002$, and PSM5, $F(1, 190) = 7.754, p = 0.003$, one-tailed. There was not a statistically significant difference in PSTs' problem-solving performance on the PSM4, $F(1, 192) = 1.830, p = 0.089$, one-tailed. The magnitude of the difference in PSTs' problem-solving performance is classified as a small effect for both the PSM3, $\eta^2 = 0.051$, and PSM5, $\eta^2 = 0.039$ (Cohen, 1988). In other words, end-of-program PSTs consistently scored better than first-year PSTs on the PSM3 and PSM5. Regarding the PSM4, fourth-year PSTs' logit scores ($M = 0.72$, $SD = 0.95$) exceeded first-year PSTs' logit scores ($M = 0.51$, $SD = 1.13$), but the difference in performance was not statistically significant.

6 | DISCUSSION

We set out to accomplish two goals in this study: (a) to examine the degree to which the PSMs, which were designed for use with grades 3–5 students, are appropriate for use with preservice teachers (PSTs), and (b) to examine PSTs ability to solve mathematical word problems. Previous validation research regarding the PSMs presents strong evidence that the PSMs are an appropriate measure of grade-level students' ability to solve problems aligned to the Common Core State Standards in Mathematics (CCSSM). We argue that evidence based on test content and evidence based on response processes, previously collected to support PSM use with grade level students (Bostic, 2018; Bostic et al., 2022), also have merit in supporting PSM use with PSTs. Specifically, that PSM items align to the CCSSM and grade-level students are able to read, interpret, and solve PSM items. We suggest if grades 3–5 students read and interpret items as intended, then PSTs ought to read and interpret items as intended as well. However, future research may also explore response processes validity evidence when using the PSMs with PSTs. For example, findings from a qualitative research study exploring how PSTs respond to PSM items may empirically support the claim that PSTs read and interpret PSM items as intended. Such future research exploring PSTs' response processes could also delve deeper into the complexity of PSM items when used with PSTs.

In this validation study, we present validity evidence based on internal structure and relationships to other variables that support using the PSMs with PSTs. We conducted a holistic examination of the PSMs' unidimensionality, and consequently present validity evidence based on internal structure to support the claim that PSMs measure a single construct. Additionally, we examined a hypothesized relationship between PSTs' performance on the PSMs based on program year. We found that end-of-program PSTs consistently outscored first-year PSTs, although the difference in performance was statistically significant for only the PSM3 and PSM5. We conclude that PSTs generally perform as expected based on a hypothesized relationship. Consequently, this validity evidence based on relationships to other variables support the claim that PST performance is related to program year; however, future research may examine this relationship in closer detail. Taken collectively, the validity evidence presented supports the use of the PSMs with PSTs as (a) formative assessments of mathematics problem-solving performance, and (b) for evaluative purposes regarding teacher preparation. This extends prior work (see Nielsen & Bostic, 2020), which indicated that PSMs for grades 6–8 were appropriate for use with preservice teachers who might teach those grade-levels. As a result of this research, mathematics teacher educators may consider using the PSMs as a formative assessment when working with preservice teachers.

PST performance on common linking items between the PSM3 and PSM4 should be noted. Students completing the PSM4 exhibited a greater probability of correctly responding to a given linking item compared to students completing the PSM3. It is particularly interesting that PSM4 is involved because ANOVA results regarding the PSM4 indicated no significant difference between PST performance. Item difficulty should be invariant across samples of PSTs; thus, differences in performance across grade level PSMs warrants further investigation.

Validation is an ongoing process (AERA et al., 2014; Kane, 2013). Additional research has potential to strengthen the validity argument for using the PSMs with PSTs. In this manuscript, we present an argument that teachers' mathematical content knowledge is important in making instructional decisions to support students' mathematical problem solving. Subsequent research of the relationship between teachers' performance on the PSMs and the instructional moves made by teachers to support students' problem solving might present stronger evidence of validity based on relationships to other variables. That is, further research may explore the relationship between teachers' PSM performance and independent measures of instructional quality. Limitations of the current study include (a) the comparison of PSTs from different cohorts, and (b) administering the

PSM online. For instance, administering the tests online may have disproportionately affected the PSM3 student subsample, which had more cases removed from data analysis compared to the PSM4 and PSM5 subsamples. Future research may employ a longitudinal design to better warrant claims about expected PST performance based on program year. We hypothesized that PST performance would be related to program year, but our findings do not necessarily support claims across all grade-levels that the teacher-preparation program had a statistically significant effect on PSTs' problem-solving performance. Future research may seek to (a) explore mathematics teacher educators' experiences using PSM scores to discuss programmatic-level decisions, and (b) examine the effect of the teacher preparation program on PSTs' PSM performance. Such future research has potential to strengthen the validity evidence based on consequences of testing and relations to other variables.

The second purpose of this study was to examine PSTs' abilities to solve mathematical problems aligned to content they may be required to teach. Our findings are analogous with those presented in Nielsen and Bostic (2020). PSTs experienced some success with solving less cognitively demanding mathematical word problems designed for the grade-levels PSTs intend to teach. In consideration of the descriptive statistics presented in Table 6, an average logit score greater than zero for all PSMs suggests PSTs experienced some success in solving problems aligned to content they may be required to teach. Lower logit values indicate less of the latent trait, or easier items to respond to. Whereas greater logit values indicate more of the latent trait, or more challenging items to respond to. For each PSM, PSTs' average logit score exceeded the average logit score of the items on the test. This implies the typical preservice teacher is more likely to respond to an item of average difficulty correctly rather than incorrectly. The probability of a test taker correctly responding to an item is 0.5 when the item measure is equal to the person measure (Embretson & Reise, 2013).

However, the average logit scores also clearly indicate that PSTs struggled to solve cognitively demanding problems, supporting the claim that the PSM items are complex for PSTs. The probability of correctly responding to a problem is less than 0.5 when the item logit-score exceeds the person logit-score. For example, the Nut Task on the PSM5 recorded a logit measure of 1.73, noticeably greater than the average PST logit score on the PSM5 of 0.83. The item, aligned to CCSSM standard 5.NBT.7, reads:

> The State Nut Company buys 22 pounds of pecans, 30 pounds of walnuts, and 31 pounds of peanuts. They sell containers of mixed nuts which contain exactly 0.5 pounds of each kind of nut. How many containers can they make?

Rasch analysis indicated the probability of the average PST to correctly respond to the Nut Task to be 0.3. The low probability of success with this item is alarming. If preservice teachers struggle to solve the problem themselves, then there must also be concern in PSTs' ability to scaffold student thinking when students experience similarly complex problems.

Student-level assessments have potential to serve as powerful tools for researchers and teacher-educators working with preservice teachers. Teacher preparation programs may collect and consider such data when evaluating program effectiveness and program improvements. Mathematics teacher educators should consider the amount of problem-solving experiences PSTs are having in their mathematics and mathematics education course work. As previously described, the magnitude of the difference in PSTs' problems solving performance was classified as a small effect. Marginal differences between first-year and end-of-program PSTs on the PSMs may indicate that PSTs are not provided enough opportunities to engage in problem solving during their coursework.

Mathematics teacher educators at the Midwest university have worked to address concerns regarding PSTs' problem-solving performance with grades 3–5 content after examining the results from the PSM data. In the last two years while these data were gathered, two new mathematics courses taught by mathematics educators have been designed and added to program requirements. One of those courses is described as an activity-based exploration of geometry and measurement concepts taught in grades PreK-5. Course content is taught through a problem-solving approach, and a goal of the course is for students to develop a deeper understanding of mathematical concepts required to teach elementary mathematics while developing proficiencies in the Standards for Mathematical Practice. The second course is described as an in-depth study of transdisciplinary learning through science, technology, engineering, the arts, and mathematics (STEAM). Emphases of the STEAM course include project based learning, and developing the content knowledge needed to plan and assess transdisciplinary learning experiences. Both courses provide PSTs with additional mathematical problem solving opportunities. Experiences engaging in rich mathematical problem-solving activities promote mathematical understanding (Lambdin, 2003), particularly mathematical content knowledge (White et al., 2013).

This study used an instrument designed to measure grades 3–5 students' mathematical problem solving performance to evaluate PSTs' ability to solve word problems

aligned to content they may be required to teach. Mathematical problem solving is a major face of mathematics education (NCTM, 2000). As indicated by the AMTE (2017) standards, novice teachers should be able to solve the problems their students solve. An important question inherent to this research is whether PSTs possess the mathematical content knowledge needed to facilitate problem solving activities with elementary students. If teachers are expected to scaffold student thinking during times of problem solving, then it is reasonable to believe that teachers can solve the problems their students experience. Therefore, teacher educators might consider validating the use of other student-level instruments to measure PSTs' knowledge, in the hopes to better prepare the teachers of tomorrow.

ACKNOWLEDGMENT

Ideas in this manuscript stem from grant-funded research by the National Science Foundation (NSF 1644214, 1720646, 1920619, 1920621). Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the National Science Foundation.

ORCID

Timothy D. Folger https://orcid.org/0000-0002-2621-343X
Jonathan Bostic https://orcid.org/0000-0003-2506-0491
Toni A. May https://orcid.org/0000-0001-7264-5607

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

Association of Mathematics Teacher Educators. (2017). *Standards for preparing teachers of mathematics*. amte.net/standards.

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

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1416898.

Bostic, J. (2018). Content validity evidence for new problem-solving measures (PSM3, PSM4, and PSM5). In T. Hodges, G. Roy & A. Tyminski (Eds.), *Proceedings for the 40th annual meeting of the North American chapter of the International Group for the Psychology of mathematics education*.

Bostic, J., Lesseig, K., Sherman, M., & Boston, M. (2021). Classroom observation and mathematics education research. *Journal of Mathematics Teacher Education*, *24*(1), 5–31.

Bostic, J., Matney, G., Folger, T., Brown, N., Evans, E., Sondergeld, T., & Stone, G. (2022). *Deepening the validity argument for the problem-solving measures 3-5*. Paper presentation by 20th annual Hawaii International Conference on Education, Waikoloa, HI.

Bostic, J., Pape, S. J., & Jacobbe, T. (2016). Encouraging sixth-grade students' problem-solving performance by teaching through problem solving. *Investigations in Mathematics Learning*, *8*(3), 30–58.

Bostic, J., Sondergeld, T., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement*, *18*(2), 151–162.

Bostic, J., & Sondergeld, T. A. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics common core. *School Science and Mathematics*, *115*(6), 281–291.

Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., DePiper, J. N., Frank, T. J., Griffin, M. J., & Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, *45*(4), 419–459.

Carney, M. B., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, *53*(4), 334–340.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Copur-Gencturk, Y. (2015). The effects of changes in mathematical knowledge on teaching: A longitudinal study of teachers' knowledge and instruction. *Journal for Research in Mathematics Education*, *46*(3), 280–330.

Curcio, F. R., & Artzt, A. F. (2003). Reflecting on teaching mathematics through problem solving. In F. Lester, Jr. (Ed.), *Teaching mathematics through problem solving* (pp. 127–142). National Council of Teachers of Mathematics.

Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950–963.

Dunekacke, S., Jenßen, L., & Blömeke, S. (2015). Effects of mathematics content knowledge on pre-school teachers' performance: A video-based assessment of perception and planning abilities in informal learning situations. *International Journal of Science and Mathematics Education*, *13*(2), 267–286.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

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

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.2307/23353796

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.

Lambdin, D. V. (2003). Benefits of teaching through problem solving. In F. Lester, Jr. (Ed.), *Teaching mathematics through problem solving* (pp. 3–14). National Council of Teachers of Mathematics.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of teachers of mathematics* (pp. 763–803). Information Age.

Linacre, J. (2012). *Winsteps (version 3.74) [computer software]*. Winsteps.com.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. *Handbook of Educational Psychology*, *2*, 287–303.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Author.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. Author.

National Governors Association & Council of Chief State School Officers. (2010). *Common core state standards mathematics*. Author.

Nielsen, M., & Bostic, J. D. (2020). Informing programmatic-level conversations on mathematics preservice teachers' problem-solving performance and experiences. *Mathematics Teacher Education and Development*, *22*(1), 33–47.

Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics*, *26*(1), 42–47.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, *67*(1), 37–58.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Danish Institute for Educational Research.

Reed, S. K. (1998). *Word problems: Research and curriculum reform*. Routledge.

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, *26*(1), 108–116. https://doi.org/10.7334/psicothema2013.260

Schoenfeld, A. H. (2011). *How we think: A theory of goal-oriented decision making and its education applications*. Routledge.

Schroeder, T. L., & Lester, F. K. (1989). Developing understanding in mathematics via problem solving. In P. Trafton (Ed.), *New directions for elementary school mathematics* (pp. 31–42). National Council of Teachers of Mathematics.

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

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205–231.

Smith, R. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, *3*, 25–40.

Varma, S. (2006). *Preliminary item statistics using point-biserial correlation and p-values.* Educational Data Systems Inc.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, *1*(3), 195–229.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse.

White, D., Donaldson, B., Hodge, A., & Ruff, A. (2013). Examining the effects of math Teachers' circles on aspects of Teachers' mathematical knowledge for teaching. *International Journal for Mathematics Teaching & Learning*. Retrieved from http://www.cimt.org.uk/fournal/white.pdf

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

Wright, B. D. (1992). Point-biserials and item fits. *Rasch Measurement Transactions*, *5*(4), 174.https://www.rasch.org/rmt/rmt54a.htm

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials*. Wide Range.

How to cite this article: Folger, T. D., Stewart, M., Bostic, J., & May, T. A. (2022). Validating the use of student-level instruments to examine preservice teachers' mathematical problem solving. *School Science and Mathematics*, *122*(8), 417–428. https://doi.org/10.1111/ssm.12558