Exploring Teachers' Pedagogical Content Knowledge of Teaching Fractions

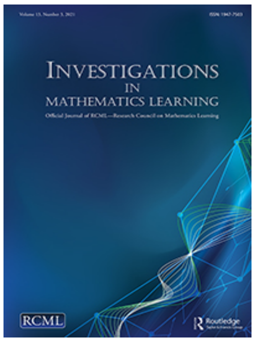
Article *in* Investigations in Mathematics Learning · September 2021

DOI: 10.1080/19477503.2021.1963145

Maryam Zolfaghari
Kent State University

Christine K Austin
Texas State University

Karl W. Kosko
Kent State University



Routledge
Taylor & Francis Group

Investigations in Mathematics Learning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiml20

Exploring Teachers' Pedagogical Content Knowledge of Teaching Fractions

Maryam Zolfaghari, C. K. Austin & K. W. Kosko

To cite this article: Maryam Zolfaghari, C. K. Austin & K. W. Kosko (2021) Exploring Teachers' Pedagogical Content Knowledge of Teaching Fractions, Investigations in Mathematics Learning, 13:3, 230-248, DOI: 10.1080/19477503.2021.1963145

To link to this article: https://doi.org/10.1080/19477503.2021.1963145

Published online: 10 Sep 2021.

INVESTIGATIONS IN MATHEMATICS LEARNING
2021, VOL. 13, NO. 3, 230–248
https://doi.org/10.1080/19477503.2021.1963145

Routledge
Taylor & Francis Group

Exploring Teachers' Pedagogical Content Knowledge of Teaching Fractions

Maryam Zolfaghari, C. K. Austin, and K. W. Kosko

Teaching, Learning & Curriculum Studies, Kent State University, Kent, Ohio, USA

ABSTRACT

The quality of teachers' mathematical knowledge for teaching (MKT) is critical for effective teaching and mathematical learning of students. However, most efforts on measuring MKT tend to focus on teachers' content knowledge (CK), with less attention to teachers' pedagogical content knowledge for teaching (PCK). This study reports on our initial efforts to develop and pilot a measure for assessing teachers' PCK for fractions. Analysis of cognitive interviews from two expert teachers combined with Rasch modeling of 85 pre-service and in-service teachers was conducted to examine validity evidence for the PCK-Fractions measure. Results provide useful validity evidence for the initial validity argument of the measure. Namely, evidence suggests differences between pre-service (PSTs) and in-service teachers' (ISTs) scores based on their professional level (junior PSTs, senior PSTs, & ISTs). Implications of this and additional validity evidence suggest a measure useful for assessing the effect of teacher education and professional experience initiatives, as well as indicators for revising this initial measure.

KEYWORDS

Elementary school; fractions; pedagogical content knowledge; validity argument

Mathematics teachers' professional knowledge plays a significant role in their effective teaching and in students' learning of mathematics (Hill et al., 2008). Such knowledge is defined by Ball et al. (2008) as mathematical knowledge of teaching (MKT) with two pragmatically distinct elements: content knowledge (CK) and pedagogical content knowledge (PCK). Although acquiring skills in both domains is essential for the quality of teaching, there are relatively few measures of teachers' PCK constructed from either a theoretical or statistical perspective (Copur-Gencturk et al., 2019; Hill et al., 2008). Copur-Gencturk et al. (2019) suggested one possible reason for undertheorizing PCK in assessment efforts is that most PCK measures are better described as measures of CK. We conjecture what Copur-Gencturk et al. (2019) observed may be due to a lack of a theoretical framing of PCK in most test development efforts. Thus, by designing an MKT measure that assesses both CK and PCK, there is a risk of unintentionally overemphasizing elements of CK and underrepresenting elements of PCK (Copur-Gencturk et al., 2019; Hill et al., 2008).

In this article, we report on an initial validity argument for a measure of teachers' (in-service and pre-service) PCK for teaching fractions to upper elementary students (grades 3–5). The PCK-Fractions measure is designed specifically to examine teachers' knowledge of children's reasoning about fractions, described by Hill et al. (2008) as knowledge of content and students (KCS). Our exclusive focus on KCS allowed us to study the nature of PCK more closely by constructing such a measure from the ground up, with an eye toward eventually addressing other domains of PCK (i.e., knowledge of content and teaching & knowledge of curriculum). Stated differently, rather than construct an overarching MKT-fractions measure including CK and PCK and given Copur-Gencturk et al.'s (2019) critique, we sought to develop a measure beginning with fewer subconstructs of PCK (i.e., KCS). Therefore, the purpose of this study is to examine initial evidence for a validity argument of an

CONTACT Maryam Zolfaghari mzolfagh@kent.edu 410_E White Hall Kent State UniversityKent, OH 44242

assessment of teachers' PCK-Fractions. We hypothesized that teachers with particular scores on the PCK-Fractions measure use specific reasoning through each item, thus providing evidence toward response processes, as well as validity for relationships to other variables.

Theoretical Framework

Many scholars have sought to research PCK, particularly in the area of mathematics education, since the construct was introduced by Shulman (1986) (e.g., Ball et al., 2008; Izsák, 2008). Perhaps the most well-known application of Shulman's construct is the MKT assessment described by the Learning Mathematics for Teaching (LMT) project (Ball et al., 2008; Hill et al., 2008). The theoretical framework described by Ball et al. (2008) includes the primary domains of CK and PCK, with each of these primary constructs including several subconstructs. While CK focuses on teachers' knowledge of doing mathematics, PCK concentrates on teachers' knowledge of pedagogy of this content, such as recognizing students' conceptions and misconceptions of mathematical topics. Measures for CK and PCK have been constructed for specific mathematical concepts such as High School geometry (Herbst & Kosko, 2014) and for multiple mathematical concepts within the same measure at both the secondary (Khakasa & Berger, 2016) and elementary level (Hill et al., 2008). Common across such studies is the finding that teachers' experience and quality of experiences are associated with teachers' PCK (Herbst & Kosko, 2014; Hill, 2010; Khakasa & Berger, 2016).

Assessments focusing exclusively on pre-service teachers' (PSTs) CK of fractions are not uncommon (Erdem, 2016; Huang et al., 2009; Izsák et al., 2019). These studies suggest PSTs demonstrate less mastery of content knowledge of fractions than what is believed needed for teaching. For example, middle school mathematical PSTs may demonstrate adequate performance of certain fraction procedures and computations, but often do not demonstrate the associated conceptual knowledge (Erdem, 2016). Similarly, Huang et al. (2009) suggested that elementary schools' PSTs' fraction procedural knowledge outweighed their conception knowledge. Contrasting the growing number of CK assessments, few studies focus on measuring teachers' PCK (Copur-Gencturk et al., 2019). Assessments that do include PCK items typically include CK items, although there are assessments that include only CK items (excluding PCK). Validation studies including PCK items suggest there is a positive correlation between PCK and CK (Depaepe et al., 2015; Kazemi & Rafiepour, 2018; Tröbst et al., 2018). However, such a correlation does not indicate causation. For example, Depaepe et al. (2015) found that secondary and elementary PSTs with different levels of CK did not show a statistically significant difference in their PCK scores. Buforn et al. (2020) provided a similar finding suggesting that elementary PSTs who were able to solve a fraction task correctly were not able to recognize mathematical aspects in students' responses. This indicates that having higher assessed levels of CK may be necessary but not a sufficient condition for higher levels of PCK (Buforn et al., 2020). Somewhat contrasting these prior two studies, Tröbst et al. (2018) found that elementary PSTs can enhance their knowledge of PCK for fractions directly, despite having a limited CK of the subject. Beyond indicating a complicated relationship between PCK and CK, results of these studies indicate that CK is not a proxy for PCK (Depaepe et al., 2015; Tröbst et al., 2018).

The inclusion of PCK items alongside CK items in test development has led to issues with validity. Copur-Gencturk et al. (2019) criticized one of the items that intended to assess teachers' PCK about ordering of fractions; "which of the following lists of fractions would be best for helping students learn to develop several different strategies for comparing fractions" (p. 487). They stated that to find the correct answer, teachers are only required to have content knowledge of the concept without being challenged with the pedagogical aspect of it (Copur-Gencturk et al., 2019). Hill et al. (2008) noticed a similar phenomenon with one PCK item stating, "a girl who was asked to count out what the 2 represents in 23, and she represented it with two checkers" (p. 391). The item was intended to assess teachers' knowledge of PCK, but during cognitive interviews some teachers used their CK to solve it instead (Hill et al., 2008). One potential reason distinguishing content in CK and PCK items is difficult is that PCK involves using one's content knowledge in conjunction with their understanding of mathematical teaching and learning (Copur-Gencturk et al., 2019). Acknowledging this, Hill et al. (2008) suggested that scholars seeking to examine this construct focus on assessing teachers' knowledge of students' mathematical conceptions and misconceptions (a particular instantiation of KCS). Additionally, Hill et al. (2008) advocated careful thought be given to the content of each individual item and how the collection of items might intersect. Following these recommendations, we sought to create an initial framework (or construct map) to characterize how individual items may focus on KCS and how these items related to one another in a broader sense.

A Hypothetical Construct Map for PCK-Fractions

To measure PCK-Fractions as a construct, we must understand what this construct involves – particularly regarding what is uniquely measuring PCK and not CK. Rather than considering a set of isolated skills, we conjectured how PCK for fractions manifests in various ways. Borrowing from recommendations by Ball and colleagues (Ball et al., 2008; Hill et al., 2008), we turned to the research literature on children's fraction reasoning to develop an initial concept map for PCK-Fractions. We cite a portion of this literature here to help describe our process for item writing. A *construct map* in the context of validity argument should not be confused with the pictorial representations often used in schooling. Rather, Wilson (2005) describes construct maps as a set of qualitative descriptors for how the construct "extends from one extreme to another" (p. 6). As a reminder to the reader, we focused specifically on KCS as a first step in conceptualizing and measuring PCK-Fractions. KCS focuses on teachers' understanding of students' mathematical thinking. Thus, we examined literature on children's reasoning with fractions (see Table 1), conjecturing that KCS for fractions may correspond with children's learning progressions for fractions (i.e., assessing children's fair sharing may be easier for teachers than assessing their part-whole reasoning). This process involved summarizing and then synthesizing the literature on children's learning progressions, and observed mathematical reasoning involving fractions (see Table 1).

Initially, children apply different real-life methods, such as one-to-one correspondence, fragmenting, and partitioning to share models (Empson, 1995). Examining children aged 3–4 years old, Hunting and Sharpley (1988) found that most children successfully shared 12 crackers among three dolls using one-to-one correspondence or the "dealing out" method. They also use their rudimentary understanding of fragmenting to fracture a cookie into halves. Similarly, Wilson et al. (2011) found that children at early ages use their informal knowledge of partitioning to share a single continuous

Table 1. Initial construct map for PCK-fractions related to children's reasoning with fractions.

Stages	Description of KCS at each stage	Example Sources
Sharing Discrete Models	Assesses children's ability to use "dealing out" method to share discrete items.	Hunting & Sharpley (1988)
Fragmenting	Assesses children's ability to use their knowledge of sharing a continuous whole into two or three parts, but these parts need not be equal.	Hackenberg et al. (2016)
Partitioning	Assesses children's ability to use their informal knowledge to share a whole equally among less than five people and exhaust the whole.	Hackenberg et al. (2016)
Equi-Partitioning	Assesses children's ability to view a partitioned whole (share a whole equally) in terms of parts within the whole and out of the whole (no limitation on parts).	Empson (1995); Hackenberg et al. (2016)
Understanding Non-Unit Fractions	Assesses children's ability to use physical representations of fractions primarily to help them to understand non-unit fractions (sequentially ordering fractions, equivalent fractions, and magnitude/ratio), then detached from concrete models and see fractions as a ratio or a number itself.	Post et al. (1985)
Understanding Improper Fractions	Assesses children's ability to construct improper fractions by applying their whole number reasoning onto unit and non-unit fractions.	Tzur (1999)

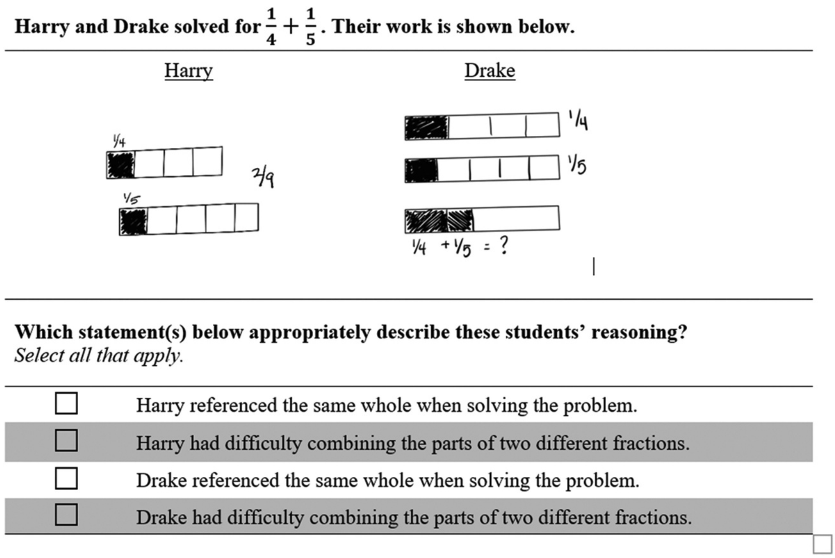


Figure 1. The example of item (item F01) for measuring PCK-fractions.

model (such as pizza) among two, four or eight people. This skill improves in sophistication when children develop part-whole reasoning. Part-whole reasoning occurs when children view a partitioned whole in terms of parts within the whole and out of the whole (Hackenberg et al., 2016). For instance, a child can see "a sandwich cut into fourths as one whole sandwich and also as a grouping of four-size pieces of a sandwich, because each scenario describes the same amount of sandwich" (Empson et al., 2020, p. 279).

To understand children's reasoning with non-unit fractions, Post et al. (1985) examined students' strategies on tasks that related to ordering fractions, equivalent fractions, and the magnitude/ratio. They found that children use physical representations of fractions primarily to help them understand non-unit fractions, then later detach from concrete models to conceptualize fractions as the ratio of two numbers. Later, Tzur (1999) found that children who were not familiar with the notion of improper fractions were able to construct improper fractions by applying their whole number concepts onto their unit and non-unit fraction knowledge (e.g., doubling $\frac{6}{11}$ to make the $\frac{12}{11}$). For these children, making an improper fraction was not simply adding or multiplying a non-unit or unit fraction, but understanding of the relationship between the possible wholes and the relation within the parts and wholes.

The hierarchy of stages presented in Table 1 represents a general overview of research from different theoretical perspectives on children's fractions reasoning. We used this hierarchy as an initial construct map for developing our measure. Specifically, we conjectured that teachers' KCS for fractions may have a parallel trajectory to children's learning progression for fractions (i.e., it is easier to assess children's part-whole reasoning than their ability to work with non-unit fractions arithmetically). In this context, we designed PCK-Fractions to examine teachers' ability to assess students' reasoning.

Development of the Measure

Following suggestions by Ball and colleagues (Ball et al., 2008; Hill et al., 2008) and guided by our initial construct map (Table 1), we designed closed-response items to measure PSTs' PCK. Closed-response items allow for quantitative analysis are less time-consuming for coding, and are typically less time intensive for participants. We used our initial construct map (Table 1), to create tasks of teaching as stems (see Table 1). *Tasks of teaching* represent scenarios of professional practice, which align with our efforts to contextualize PCK items (Ball et al., 2008). For example, the item shown in Figure 1 represents two different children's reasoning about the concept of the whole in the context of adding fractions. The task of teaching involved in this example includes assessing children's partitioning and equipartitioning (Table 1).

After initial writing and revision of items, we engaged in a process of collecting initial validity evidence to establish a validity argument for our preliminary construct map (test content, response processes, internal structure, relationships to other variables). Collection and analysis of this evidence aligns with the purpose of this study, which is to design a measure of PCK for fractions and construct an initial validity argument for that measure. To do this, we sought to answer the following research questions:

(1) Does the PCK-Fractions measure assess the PCK domain?
(2) How does the preliminary Construct Map for PCK-Fractions align with validity evidence collected?
(3) How does the PCK-Fractions measure distinguish between teachers' scores based on their progress through teacher education (junior PSTs, senior PSTs, & in-service teachers)?

In posing these research questions and stating the above purpose, it is worth noting that we considered these efforts as the *first step among many* in developing our PCK-Fractions measure. As noted by Krupa et al. (2019), "most validity arguments are the result of numerous studies … It is the accumulation of evidence from various sources for claims that builds a coherent validity argument, not a single study or report" (p. 11). In this vein, we position this article as providing initial evidence to establish a validity argument. Simultaneously, we also view the work presented here as identifying what additional validity evidence is needed and how the PCK-Fractions measure may be improved.

Methods

Sample

In fall 2019, we surveyed a total of 85 participants which included 58 PSTs and 27 ISTs. PSTs were enrolled in a teacher education program in a university located in the Midwestern U.S., and ISTs were recruited from the same geographic area with at least 3 years of teaching experience in upper elementary mathematics (grades 3–5). Additionally, our sample included two elementary mathematics coaches (i.e., an expertise teacher in math and pedagogy who provides professional development to classroom teachers), Zoe and Wade, who participated in cognitive interviews prior to construction of the piloted measure.

Excluding the two cognitive interview participants, most participants self-identified as female (91.7%), and the remaining identified as male (8.3%). The PST participants were enrolled in one of the two majors for teacher licensure: early childhood (n = 47) and middle childhood education (n = 11). At the time of data collection, early childhood licensure included preschool through third grade, with an optional fourth- & fifth-grade endorsement. Middle childhood licensure included grades 4–9. The PST participants were recruited from a mathematics methods class via face-to-face and emailed solicitation. Each licensure program included two mathematics methods courses with the first occurring in the second semester of participants' junior year (i.e., 3rd of 4-year program) and the second occurring in the first semester of participants' senior year(i.e., 4th of 4-year program). PST participants were solicited from both sets of courses across both programs (29 juniors; 29 seniors). ISTs were recruited via e-mail and completed the survey virtually. ISTs reported an average of 17.15 years of experience (*Range* = 4–32 years). Participation of ISTs was limited to those with at least 3 years of experience given such experience often distinguishes between novice and experienced teachers (Herbst & Kosko, 2014).

Measure

Earlier in this article, we described the development of the measure for PCK-Fractions. The PCK-Fractions assessment included 15 questions: 9 multiple-choice and 5 multiple-response. *Multiple-choice items* included a stem where only one choice may be selected (i.e., select the best choice). *Multiple-response items* included the same stem, but each option represented an independent item (i.e., select all that apply). Specifically, each multiple-response option was treated as an item for coding following the suggestions from Linacre (2021). Raw responses were dichotomized (0 = incorrect; 1 = correct). So, the 15 questions represented 30 distinct items for item analysis (see Appendix for additional items).

Analysis

Analysis focused on four forms of validity described in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014): test content, response processes, internal structure, and relationships to other variables (see Table 2). Validity evidence for *test content* focuses on how an assessment represents the content it is designed to assess, and whether the assessment scores can be interpreted in line with the assessment's intended purpose (AERA et al., 2014). The intended purpose for our PCK measure is to examine teachers' level of understanding in assessing students' reasoning with different professional experiences. This purpose implies an intended use of PCK-Fractions scores to examine the effect of different professional experiences (such as years of experience, teacher education, etc.) and not to evaluate individual teachers. The intended purpose and use of the assessment, as well as the construct map (see Table 1), were considered as evidence sources for test content. Additionally, we used cognitive interviews as another primary evidence for the test content. Cognitive interviews with expert teachers and psychometric data both allowed for analysis of whether the items assess PCK (and not CK) and provided initial evidence regarding the validity of the hypothetical construct map.

Validity evidence of *response processes* focuses on whether participants' responses to items align with the intended theoretical design of the item. We used evidence from cognitive interview data as a primary source for response processes, as well as item hierarchy provided by our Wright map (see description of Rasch analysis). Validity of *internal structure* addresses "the degree to which the relationships among test items and test components conform to the construct" (AERA et al., 2014, p. 16). In this study, we used indicators of unidimensionality estimated through Rasch modeling for this purpose. Validity of *internal structure* included investigating point-biserial correlations, in which performance on an item is compared to the overall test performance (Crocker & Algina, 2006), as well as item fit and other indicators of reliability (item and person). Validity for *Relations to Other Variables* refers to the relationship between assessment scores and other external variables such as categorial variables (AERAet al., 2014; Bostic et al., 2017). "Categorical variables, including membership variables, become relevant when the theory underlying the proposed test use suggested that group differences should be present or absent if a proposed test score interpretation is to be supported" (AERA et al., 2014, p. 16).

We used one-way ANOVA as an indicator of teachers' PCK scores to understand the relationship between their professional experience and PCK as initial evidence in this regard. Moreover, we used indicators of reliability of items and persons (participants) as initial evidence for this form of validity to establish the measure's capacity for distinguishing between different (sub)groups' scores. Although development of new measures requires integration of various sources and types of evidence to construct a validity argument, not all possible sources or types of evidence need be present in a single study (AERA et al., 2014). The validity evidence presented here is considered only a portion (i.e., the *first* portion) of a validity argument developed across multiple studies.

Cognitive Interviews

Cognitive interviews were employed as a primary means to assess whether participants' responses aligned with the intended purpose and theoretical design of each item (response processes). Data also were used to examine whether the content of the items was aligned with the PCK domain and not CK. Data were obtained through interviewing expert elementary mathematics teachers, Zoe and Wade. We targeted expert teachers because their knowledge "is not only more than novice teachers but also this knowledge is greatly structured" (Kazemi & Rafiepour, 2018, p. 750). So, by focusing only on expert teachers at this initial stage, we sought to focus on whether pedagogy, and not the content, was at the forefront of reasoning through each item. Following guidelines provided by Karabenick et al. (2007), the cognitive interviews included three leading questions for each assessment item: 1) What was the question asking? 2) What is your answer? and 3) Why did you choose that answer? This method of probing allows for *think-aloud* responses as participants complete each item (Karabenick et al., 2007). Analysis of responses consisted of examining whether participants' rationales for responses aligned with our item design (i.e., they answered based on their understanding of students' mathematical reasoning).

Rasch Modeling

We used Rasch modeling to examine validity evidence for internal structure and response process. The choice of Rasch modeling over a classical approach was due to Rasch being able to transform raw data, which is typically ordinal by nature, into continuous data using a logistic approach (Bond & Fox, 2015). The Rasch approach uses patterns in item responses from individuals, and patterns in individual responses across items to create a logistic model (Bond & Fox, 2015; Crocker & Algina, 2006). The logistic model estimates participants' ability level for the measured latent construct using test statistic theta (θ), as well as each item's difficulty level using test statistic delta (δ). This is particularly useful in soliciting validity evidence toward internal structure and response processes, since it assesses how the items align with the measured construct by examining the relationship between the test items and the responses of the participants. This relationship is visualized through a Wright Map (Figure 3), which juxtaposes participants' theta statistics with items' delta statistics on a vertical logistic scale (Bond & Fox, 2015; Wilson, 2005). This allows for visual analysis of the spread of item difficulty while simultaneously examining the spread of participants' assessed ability on the measure. Items' ordinal arrangement on the Wright map can be compared to the construct map used to design items (Figure 3). This in turn can be used to either revise the measure or the theory (Wilson, 2005). Thus, the Wright Map and accompanying statistics provide useful evidence toward response processes.

The initial process of Rasch modeling requires the following basic requirements to be met which include assessing for unidimensionality, local independence, and equal discrimination (Bond & Fox, 2015; Wright, 1991). For clarity, unidimensionality may refer to a measure assessing only one aspect of a construct (i.e., PSTs and ISTs' level of PCK). In this article, we examined unidimensionality by evaluating item and person reliability, Rasch principal component analysis (PCA), infit and outfit, and the Wright map. Checking for local independence requires checking for the fit statistics for our items to assure that they are not dependent on each other. Additionally, equal discrimination of our items is checked through the item difficulty, a logit score above 1.0 suggests that the item is able to differentiate between high and low ability participants (Linacre, 2021). Additionally, checking the Rasch requirements helps us to gather validity evidence toward PCK-Fractions.

Our evidence toward internal structure validity was collected in various ways including item and person reliability, fit statistics, and unidimensionality of the construct (AERA et al., 2014). Variance and fit statistics associated with individuals (person reliability index) or with items (item reliability index) can be teased apart and/or compared (Bond & Fox, 2015). For reference, person reliability is the replication of person order when using the same sample on "a parallel set of items measuring the same construct" (Bond & Fox, 2015, p. 49). Item reliability is the replication of item order "if the same items were given to another same-sized sample of persons who [behave] in the same way" (Bond & Fox, 2015, p. 49). Thus, we examined two types of fit statistics: infit (weighted) and outfit (unweighted) which estimate the responses given the predicted response (Bond & Fox, 2015). A low-test reliability coefficient, such as Cronbach's alpha, may be due either to issues with item design, sample, or both (Crocker & Algina, 2006). Rasch allows for separate reliability and fit measures, thereby narrowing down whether and where potential issues with reliability reside.

We also incorporated Rasch principal component analysis (PCA) of the residuals to check for unidimensionality (Bond & Fox, 2015). This analysis determines how much of the variation is explained in the model. For this to be valid, there are four basic assumptions that should be met: 1) The amount of variability explained by the total number of residuals should be 4 times greater than the variability explained by the first contrast, 2) the amount of variability explained by the total residual model should be greater than 50%, 3) the eigenvalues of the contrast should be less than 3.0, and 4) the variability of the first contrast should be less than 5% (Linacre, 2021; Wright & Stone, 2004). If any assumption is violated, investigation of items and/or persons is warranted. The use of PCA along with item and person reliabilities and infit/outfit statistics allows for a more cohesive story exclaiming if the construct is assessing the single attribute of PSTs and ISTs PCK of fractions – whereas PCK in this article focuses specifically on knowledge of content and students (KCS).

ANOVA

We used a one-way between subjects' analysis of variance (ANOVA) to examine differences in teachers' experience levels related to teacher education. This examination aligned with our third research question as well as providing evidence for a validity of relation to other variables (i.e., correlation to external validity of level of experiences). In other words, we used the ANOVA analysis as a way of examining for differences in PCK scores as related to different levels or amounts of experience (PST juniors, PST seniors, experienced ISTs). Besides, we compared PCK theta Rasch model scores of PST juniors, PST seniors, and ISTs with at least 3 years of experience (Author, 2014; Hill, 2010).

Results

Cognitive Interviews

As an initial phase of validation of response processes and test content, two expert elementary mathematics teachers (Zoe & Wade) were recruited to partake in cognitive interviews for the PCK-Fractions assessment. Feedback from the two cognitive interviews allowed us to eliminate two items due to interpretability issues, five items needing revision, eight items unchanged, and five items that were set aside due to assessing similar content. The two eliminated items were replaced with two lower difficulty items bringing the overall PCK-Fractions assessment to 15 questions (30 items). Figure 2 illustrates an item that was meant to be easier but led to difficulty for our experts. One participant, Zoe, struggled with the item by stating that, "I can understand the representation but . . . I'm second guessing myself. I see what he did and understand it but, I'm not convinced it's right . . . I don't know if it's the multiplication that's incorrect or the picture." Similarly, Wade also questioned his reasoning by thinking aloud "well he counted all the parts of both of the fractions . . . that's what he did. He didn't miscount anything." Although the item was more difficult than we anticipated, evidence from the cognitive interviews does suggest that Zoe and Wade were focused primarily on student reasoning and deciphering how the child arrived at their answer. Specifically, the attention to mathematics in the items was focused on the child's mathematics and not that of our expert teachers. Thus, cognitive interview data on this and other items consistently indicated that the items were focusing on the PCK construct. However, the evidence suggested that the hypothetical construct map we conjectured may

Table 2. Summary of validity propositions and accompanying evidence.

Validity Type	Claim(s)	Primary Evidence
Test Content	• The measure assesses PCK (not CK) of fractions. • Teachers with more professional experience(s) have higher PCK-Fractions.	• Purpose & Intended Use of Measure • Construct Map • Cognitive Interviews
Response Processes	• The measure assesses PCK (not CK) of fractions. • PCK-Fractions has different degrees of sophistication.	• Cognitive Interviews • Construct Key Map & Wright Map
Internal Structure	• PCK-Fractions is a unidimensional construct.	• Infit & Outfit Statistics • PCA of Standardized Residuals • Construct Key Map & Wright Map
Relation to other variables (level of experiences)	• The measure can assess the effect of professional experience across different groups/subgroups of teachers.	• Item Separation & Reliability • Person Separation & Reliability • One-way ANOVA of PCK-Fractions scores

need to be modified. Since the items were generally well interpreted, with five items needing minor revisions, we decided to move forward with a pilot sample to better understand the nature of the items and construct.

Rasch Modeling

We implemented Rasch modeling to collect certain forms of validity evidence. Recall that to incorporate this approach, certain requirements should be addressed. An issue of Rasch model fit is reliant on three requirements: unidimensionality, equal item discrimination, and local item independence. These general requirements are instilled "in order to achieve invariant interval-level measurement" (Bond & Fox, 2015, p. 265) that ensures our data fit the model. These requirements are addressed by examining mean-square fit statistics (i.e., infit and outfit), point biserial correlations, and Rasch PCA. Our findings suggest that the Rasch model of the measure of PCK-Fractions has sufficient item reliability of 0.90 (acceptable if near or above 0.90) and a sufficient item separation reliability of 3.0 (Duncan et al., 2003; Linacre, 2021). An item separation index above a "2.0 represents a good level of separation" of the items and can differentiate between high, average, and low levels of item difficulty (Duncan et al., 2003, p. 953). This suggests that the measure can distinguish between more difficult and easier items. In other words, this insinuates that, when administering PCK-Fractions to other samples, the item difficulty should remain relatively constant. Thus, from our item reliability, we can expect that item F14 (Figure 2) would still be the most difficult item for a new set of participants, and F20d would be the easiest. By contrast, person reliability is considered acceptable if near or above .80. However, this administration of the PCK-Fractions did not have sufficient person reliability (0.41). Bond and Heene (2020) suggest two predominant causes of low person reliability: 1) too few items spread across a range of difficulty and/or 2) too little spread of participants with varying theta scores (ability). These possibilities can be examined through the Wright map in Figure 3. As illustrated in Figure 3, several items are evenly distributed between −1.00 and 2.00 logits, with a potential need for additional items targeting below −1.00 logits. However, a more likely cause for the subpar person reliability is the negative skew in participants' scores, as 75.3% of participants have a score above 0.00, or average ability (M = 0.37, SD = 0.64). This implies that our participants' measured abilities were much higher than we originally anticipated, thus suggesting a need for a larger range in measurable ability (i.e., first and second year PSTs) or in item difficulty (i.e., more difficult items).

Table 3. Revised item analysis statistics.

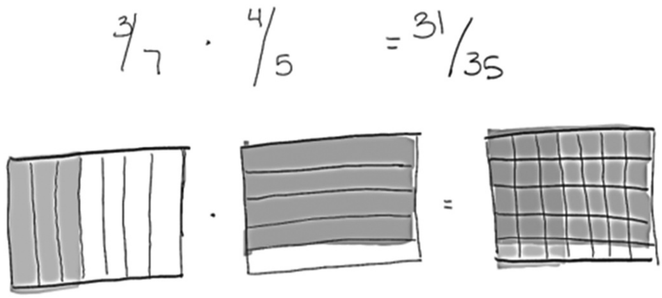
			Infit		Outfit	Point-Biserial Correlations	
Item	Item Difficulty	SE	Mean Square	*Z*	Mean Square	*Z*	
F01a	0.12	0.23	1.01	0.20	0.98	−0.20	0.28
F01c	0.66	0.23	0.94	−0.90	0.94	−0.80	0.38
F03	0.46	0.23	1.05	0.80	1.08	1.10	0.20
F05a	0.12	0.23	0.95	−0.80	0.94	−0.80	0.37
F05b	−0.39	0.24	1.02	0.20	1.02	0.20	0.24
F05c	0.39	0.23	1.04	0.70	1.04	0.60	0.22
F07	−0.84	0.27	1.01	0.10	0.95	−0.30	0.25
F08a	−0.33	0.24	1.02	0.20	1.09	0.80	0.21
F08b	0.31	0.24	1.10	1.70	1.15	1.90	0.10
F08c	−0.17	0.24	1.05	0.60	1.11	1.10	0.16
F08d	−0.86	0.28	0.94	−0.30	0.88	−0.60	0.34
F09	−0.45	0.25	1.10	1.10	1.10	0.80	0.11
F10	−0.04	0.24	0.90	−1.40	0.87	−1.40	0.43
F11	1.30	0.25	1.08	0.70	1.06	0.50	0.16
F12	−0.01	0.24	1.02	0.30	1.09	1.00	0.22
F14	1.78	0.28	1.05	0.40	1.09	0.50	0.16
F15	0.83	0.24	0.91	−1.10	0.90	−1.10	0.42
F16a	−0.64	0.25	1.02	0.20	0.99	0.00	0.23
F16b	0.18	0.23	1.07	1.20	1.06	0.80	0.18
F16c	0.18	0.23	1.00	0.00	0.97	−0.40	0.30
F20a	−0.64	0.25	1.04	0.40	1.03	0.30	0.20
F20b	−0.64	0.25	0.92	−0.70	0.85	−1.00	0.40
F20c	0.18	0.23	0.84	−2.70	0.81	−2.60	0.54
F20d	−2.75	0.52	0.94	0.00	0.64	−0.50	0.28
F20e	1.18	0.25	1.00	0.00	0.94	−0.40	0.30
F20f	−0.57	0.25	0.98	−0.10	0.98	−0.10	0.28
F21	1.09	0.25	0.90	−1.10	0.92	−0.60	0.42
F22	−0.45	0.25	1.10	1.10	1.19	1.50	0.08
Mean	0.00	0.25	1.00	0.00	0.99	0.00	0.27

Examining the above issue further, an analysis of global fit suggests that the data does not have a high degree of misfit ($\chi^2 = 2886.28, df = 2883, p = 0.48$). This suggests that the items on the PCK-Fractions construct and our sample ISTs and PSTs performance on the construct follow more of a predictive pattern for the Rasch model (Bond & Fox, 2015). According to Linacre (2021), our data not having a high degree of misfit may be due to our small sample size. Further, although person reliability was found to be low, the average mean-square person infit (*MNSQ* = 1.00, *Z* = 0.00) and outfit (*MNSQ* = 0.99, *Z* = 0.00) were at Rasch model expectations. These findings suggest that our ISTs and PSTs responded to the items as predicted by the Rasch model. Average fit indices for PCK-Fractions items also met model expectations for infit (*MNSQ* = 1.00, *Z* = 0.00) and outfit (*MNSQ* = 0.99, *Z* = 0.00) statistics. This culmination of evidence indicates that PCK-Fractions items perform well, but the lack of range in our participants' measured ability (i.e., a low person reliability) warrants a need for a wider range of ability within our sample. Such additional participants could include PSTs enrolled earlier within the education program or non-education major people.

Initial validity evidence toward internal structure includes item and person reliability described in the preceding paragraphs. It also includes an analysis of infit and outfit statistics (see Table 3). Infit and outfit statistics contribute toward evidence of a unidimensional construct. Our item-level infit statistics generally fall between the acceptable range of 0.75 to 1.33, indicating that our items fit within the predicted Rasch model (Bond & Fox, 2015).[1] For us, this means that our participants' performance on PCK-Fractions aligned with how the item should perform (based on difficulty) as predicted by the Rasch model. Typically, an infit statistic less than 0.75 indicates a potential overfit and could signify a lack of item independence.

[1]Our person fit statistics was also within the acceptable range. Our infit statistics ranged from 0.79 to 1.39 (M = 1.00, SD = 0.14). Only one person had an infit score of 1.39 with the next highest of 1.29.

Derek was asked to find the product of $\frac{3}{7}$ and $\frac{4}{5}$.
His work is shown below.



Which of the following statements best describes his work?	
①	Derek applied the standard algorithm inappropriately.
②	Derek applied the standard algorithm appropriately.
③	Derek created an appropriate representation and miscounted the parts of both fractions.
④	Derek created an appropriate representation and counted all parts of both fractions.

Figure 2. Example of an item (F14) that was more difficult than intended.

Given the use of multiple-response items, we examined evidence for such overfit but found none. However, we also examined items' point-biserial coefficients and found two items (F01b & F01d) had negative (–0.17) and 0.00 correlation coefficients, respectively.

Both items had similar phrasing, which prompted revision following cognitive interviews. However, the point-biserial coefficients suggest these items were not revised sufficiently, and they were removed. This resulted in increased point-biserial correlations across items, as well as an increased person reliability measure (0.51). The revised item statistics are presented in Table 3. Removal of F01b and F01d ensured our requirement for equal item discriminations for the Rasch model was sufficient. Notably, there are items with less than ideal point-biserial coefficients that we retained. However, given the stage of measure development (pilot) the items were retained for the present analysis to collect validity evidence for future steps in measure development. Additionally, we did not observe evidence in cognitive interviews to justify their removal.[2]

Additional evidence of unidimensionality and, therefore, evidence toward internal structure validity, comes from Rasch principal component analysis (PCA) (Bond & Fox, 2015; Wright & Stone, 2004). The first contrast identified explained 8.6% of the variance, with an eigenvalue of 2.84. Regarding Rasch PCA assumptions, the amount of variability explained (82.8%) was 10.1 times more than the variability of the first component (8.4%), exceeding the recommendation that such variance explained by the factor analysis is greater than 50% (Linacre, 2021; Wright & Stone, 2004). An eigenvalue less than 2.0 likely indicates random noise, while a value above 3.0 suggests potential for multidimensionality (Bond & Fox, 2015). Eigenvalues between 2.0 and 3.0, as observed here, still warrant investigation, but upon further investigation, the items loading on this potential factor demonstrated no obvious patterns and so no further action was taken. Although the last assumption

[2]Such items will be examined in future iterations of the PCK-Fractions measure, including interview data from PSTs.

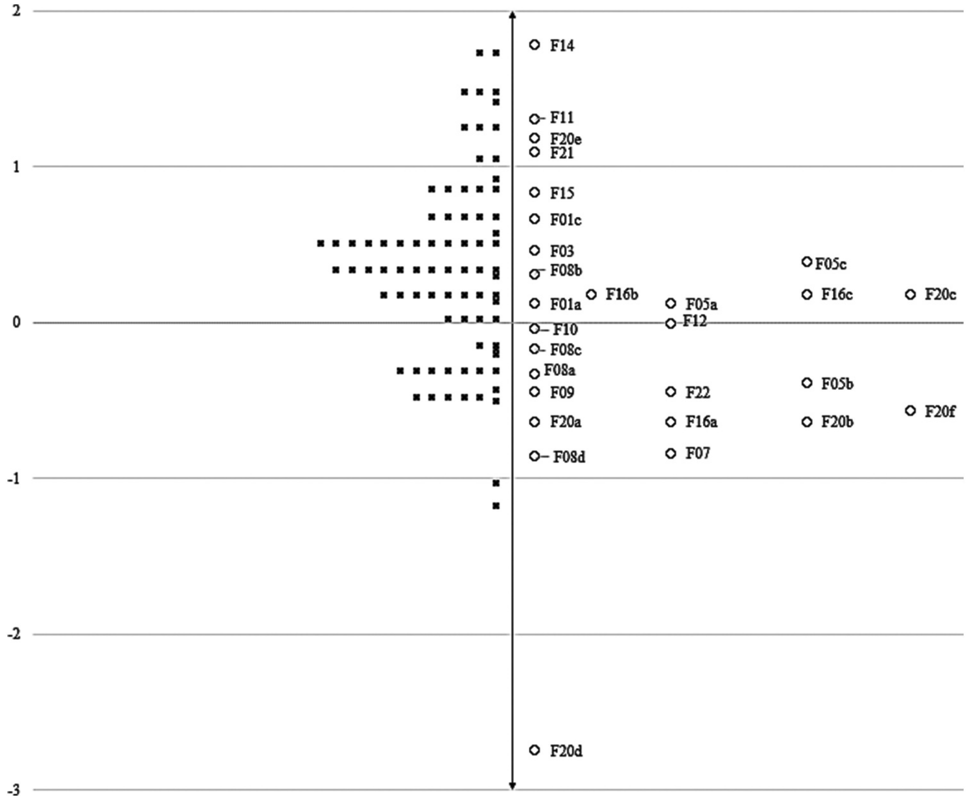


Figure 3. Wright map for PCK-fractions.

of Rasch PCA was not fully met, since variability of 8.6% is somewhat larger than the suggested less than 5% criteria, Linacre (2021) suggested that the difficulty of some items may impact the variances along with the lack of range in person PCK ability. As noted, 75.3% of our participants scored above average ($\theta \geq 0.00$) on the PCK-Fractions measure (Figure 3) implying negative skew of assessed PCK. It is also evident in Figure 3 that all items, except one (F20d), have a logit score above – 1.00, which suggests that our items are more difficult than expected. The addition of Rasch PCA to the examination of item fit via infit statistics (Bond & Fox, 2015; Wright & Stone, 2004), suggests that the items of PCK-Fractions measure the same unidimensional construct. This supports our conjecture that PCK-Fractions is assessing PCK of ISTs and PSTs as intended due to unidimensionality being met.

The next step in item analysis was to examine the Wright map (see Figure 3). The Wright map helps contribute validity evidence for internal structure and response processes. Recall that the construct map presented earlier in the article hypothesized that how children generally develop fractional reasoning would correspond to the trajectory for how teachers learn to assess such reasoning (see Table 1). Evidence from the Rasch model did not support this hypothesis as envisioned. Rather, delta statistics provided evidence of a hierarchy that corresponded less to the fraction concept addressed within a question/item and more to the observable actions of children being assessed in each task of teaching.

In our initial design of items, we hypothesized that the difficulty of assessing children's fractions reasoning would correspond with the sequence certain fraction concepts are typically learned. However, the student actions being assessed by the test takers appeared to be a better explainer of why certain items have different difficulties than predicted. Consider the four items nested in question F08 (Figure 4). We hypothesized that F08a would have the lowest delta statistic of the four, since children demonstrate this form of reasoning (using whole-number reasoning in a fraction context) prior to developing the concept of a fraction. This should be followed by F08c, F08b, and F08d, respectively. Evidence from the Rasch modeling suggests that assessing whether a child is using whole-number reasoning in a fraction context is more difficult than assessing whether a child has partitioned a visual whole into equal parts.

Comparing F08d with items of similar difficulty, it appears such items asked participants to assess children's creation or use of fractional parts. Items like F08b, while appearing to ask for the same assessment, also include a layer of difficulty with coordinating nonstandard versions of fractions (in this case, assessing a child's recognition that four $\frac{2}{3}$s, or $\frac{8}{12}$, are the same as $\frac{2}{3}$). Thus, we reexamined items with similar delta statistics to gauge whether, and to what degree, the same assessment requests were made of participants. We found that items with delta statistics below −.50 tasked participants to assess children's creation and/or use of fractional parts (i.e., F08d). Items approximately between −.50 and .00 asked participants to assess children's coordination of parts with and of the whole (i.e., F08a & F08c). Items with delta statistics approximately between .00 and .50 tasked participants with assessing children's creation and use of non-unit fractions (i.e., multiple iterations of $\frac{2}{3}$ as equivalent to $\frac{2}{3}$ in F08b). Items with delta statistics higher than .50 tasked participants with assessing children's coordination of non-unit fractions with the whole. For example, item F11 tasked participants with assessing a students' comparison of $\frac{6}{7}$ and $\frac{7}{8}$.

ANOVA

Results from ANOVA provided validity evidence toward relationships to other variables (i.e., level of experiences) and addressed our third research question. Specifically, we would expect teachers with more professional experience with assessing children's fraction reasoning would have higher PCK-Fractions scores. Results from ANOVA found a statistically significant difference in PCK scores for PST juniors, PST seniors, and in-service teachers (ISTs) [$F(2,78) = 5.616$, $p = 0.005$]. A Tukey HSD post hoc analysis indicated a statistically significant ($p = 0.040$) difference in scores between PST juniors ($M = 0.12$, $SD = 0.47$) compared to PST seniors ($M = 0.46$, $SD = 0.51$), as well as statistically significant ($p = 0.006$) between PST juniors and ISTs ($M = 0.58$, $SD = 0.56$). This indicates a measurable difference in PCK-Fractions scores for PSTs with an additional methods course and accompanying experiences (student teaching) in the teacher education program participants were sampled. In contrast, there was no statistically significant difference observed between PST seniors' scores and ISTs' ($p = 0.692$). This implies that, although ISTs' scores were higher than PST seniors', the difference was not large enough to be statistically significant. This may be due to the current sample size and range of participants and should not be considered as generalizable to the larger population of PSTs and ISTs. Rather, Linacre (1994) notes that sample sizes similar to this study (n = 81) are able to estimate participant scores within ±0.50 logits. This level of precision could account for the ANOVA results observed here. Alternatively, the results could also be due to either the quality of the teacher education program or the quality of professional experiences of the ISTs in the sample. In all, these findings provide preliminary evidence that the PCK-Fractions measure can detect some differences due to teachers' level of developmental experiences, such as progress within a teacher licensure program. To detect differences in their developmental experiences, such as those examined here, Linacre (1994) suggests including samples of at least 250 to have such precision. Results presented here suggest there is potential for the measure to detect such differences, but future study is necessary to verify such potential.

Discussion

Contrasting prior approaches that developed PCK measures alongside CK measures, we sought to develop items exclusively focused on the PCK domain, with particular attention to KCS as an initial effort to item design. We hypothesized this approach would allow for a more nuanced understanding

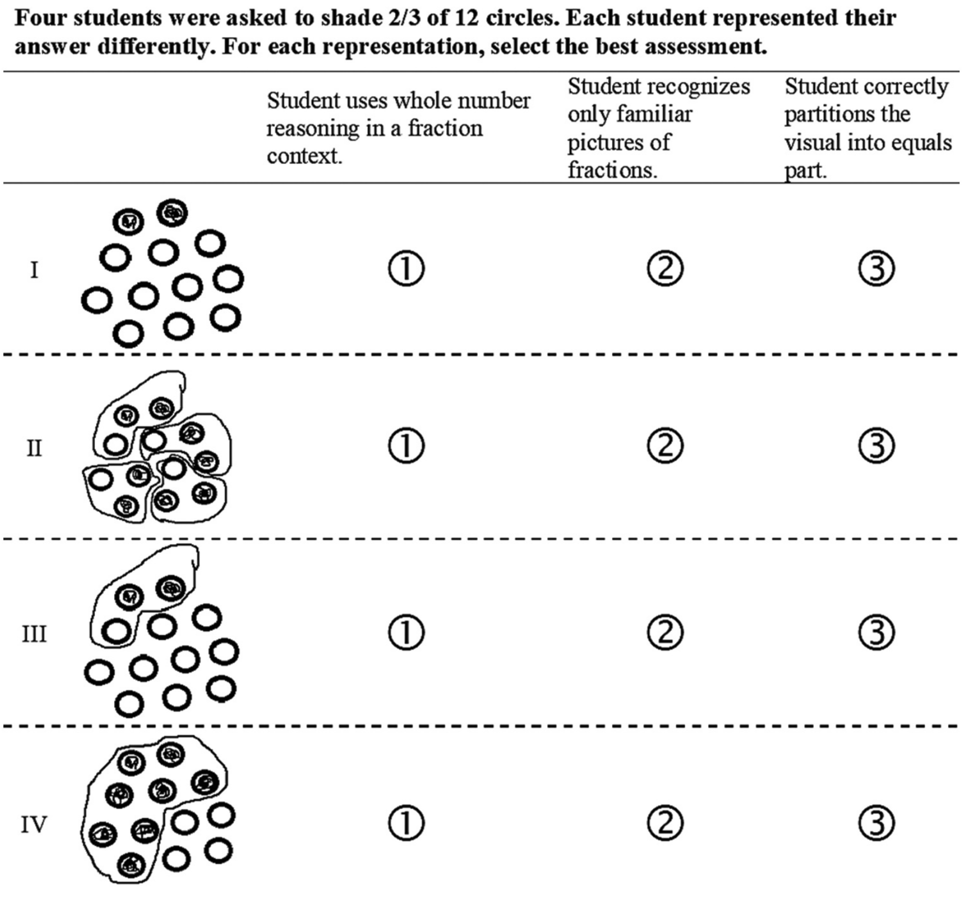


Figure 4. Top to bottom, F08a (δ = −.31), F08b (δ = .32), F08c (δ = −.16), F08d (δ = −.85).

and assessment of PCK for fractions, while reducing the risk of including CK items masquerading as PCK items (Copur-Gencturk et al., 2019; Hill et al., 2008). Such an effort required some theoretical conjecture of the nature of PCK as a domain, given the lack of scholarship in this area. Similar to Hill et al. (2008), "we attempted to build the ship while sailing it – by writing items to help push our conceptualization and definition forward" (p. 396). Such an endeavor necessitates a certain amount of failure, but with the proposition that the rewards outweigh such risks.

Regarding our initial conjectures on the nature of PCK, we hypothesized that item difficulty would mimic the trajectories which students tend to develop fractions reasoning (see Table 1). Although validity evidence supports our intent of items assessing the KCS subconstruct of PCK, findings illustrated in the Wright map (see Figure 3) and evidence from cognitive interviews suggests our initial construct map was incorrect (see Table 4). The descriptions of each level in the revised construct map (see Table 4) appear similar to those shown in Table 1. However, findings from the present study suggest that the teachers' actions being assessed at each level do not necessarily correspond with the sequence of content learned by children. The differences in difficulty for items from the F08 stem (see Figure 4) serve as a prime example of this phenomenon. To the best of our knowledge, no complete learning trajectory for PCK of fractions currently exists. However, there are recent, albeit isolated, findings that provide some preliminary support for the revised construct map presented here. For example, Tyminski et al. (2020) found that PSTs had a more difficult time analyzing one student's area model where they were asked to solve ¼ of 8 and divided each item into fourths before adding all fourths. Such reasoning would fall under Level 4 in our revised construct map (Table 4). Although alignment of such findings is helpful, there are relatively few studies that examine or describe teachers' difficulty with particular aspects of PCK. The construct map presented here provides a reference for others engaging in such work, but one that will need revision, evaluation, and expansion as scholars continue to examine teachers' PCK for fractions.

The revised construct map (see Table 4) serves as a useful starting point for study of teachers' PCK for fractions. Prior to this study, such an empirically based starting point was unavailable. However, we admit the limitations and emphasize that this construct map needs to be empirically validated and updated through quantitative and qualitative study. Regarding the PCK-Fractions measure, there is a need to revise item wording to better align with the revised construct map. Specifically, items written for the original construct map included options and particular phrasings that aligned with the hypothesized sequence and adjustment is needed to improve the items in this regard. Other research could also examine the development of PCK longitudinally through either quantitative or qualitative means. Further, evidence suggests PCK is correlated with PSTs' CK of fractions (Depaepe et al., 2015; Kazemi & Rafiepour, 2018). Yet, a proper understanding of how CK facilitates the development of more advanced PCK is lacking. Similarly, there is a need to better understand how PCK develops across various mathematical concepts and grade-band foci for educators.

Toward a More Complete Validity Argument

Despite writing items for a measure based on a flawed, albeit research-informed, initial construct map, the PCK-Fractions measure performed remarkably well for an initial pilot. Rather, if "we attempted to build the ship while sailing it" (Hill et al., 2008, p. 396), we managed not to sink, and developed an initial validity argument that can be expanded upon with a basis of empirical evidence. Recall that the intended purpose for the PCK-Fractions measure is to assess professional knowledge of teachers with different professional experiences. As such, the intended use of the measure is to evaluate the effectiveness of various experiences that are intended to contribute to improving teachers' PCK. Validity evidence presented in this article provides initial support toward this purpose and use statements. However, we have and continue to stress validity evidence and validity arguments are best constructed over a series of several studies (Krupa et al., 2019). Therefore, results and findings in this article provide support for an initial validity argument while simultaneously indicating next steps for improving the measure and constructing the validity argument.

Analysis of cognitive interviews provided evidence toward the response processes and test content of the measure aligning with the PCK subconstruct of KCS. Coupled with the results from the Rasch analysis, this set of validity evidence indicated our initial construct map was flawed. In essence, we needed the psychometric data to fully understand and make use of evidence that emerged in the cognitive interviews. Given the revised construct map (Table 4), and the need for new items to assess lower PCK, an additional step in the validation of the PCK-Fractions measure is to review the structure and wording of all items for better correlation with the revised construct map. Additionally, interviewing PSTs at different stages in their teacher education about their responses may provide a better understanding of the nature of their PCK and how it develops. Given the need for future revision, we provide a sample of items in the Appendix, but not the entire instrument. We believe this allows for transparency regarding the strengths and current limitations of the measure following its *initial* pilot.

ANOVA results were used as evidence related to other variables (i.e., level of experience). This preliminary evidence suggests that the PCK-Fractions measure may detect differences due to professional experiences (i.e., participation in a teacher education program). Such evidence provides the first indication of relationships to other variables, and additional evidence is

Table 4. Revised construct map for PCK-fractions.

Level	Description	Items[a]
Level 1	Assess children's creation and/or use of fractional parts.	F07, F08d
Level 2	Assess children's coordination of parts and of the whole.	F08a, F08c, F09
Level 3	Assess children's creation and use of non-unit fractions & comparison of fractions.	F01a, F08b
Level 4	Assess children's coordination of non-unit fractions with the whole & comparison of fractions and wholes.	F01c, F14, F21

[a]These are example items provided in the Appendix.

necessary to better evaluate PCK-Fractions' capacity in this regard. Such evidence may include both cross-sectional and/or longitudinal data. Second, evidence from Rasch modeling indicated the current measure lacks items with lower difficulty, and the current sample included too few such individuals. Thus, in addition to sampling participants, we anticipate would have low PCK scores, next steps should include writing and piloting items focusing on assessing lower PCK.

Rasch analysis provided validity evidence for internal structure, as results from Rasch modeling indicated a need for items of less difficulty and participants with lower levels of PCK for fractions. This latter finding, coupled with the ANOVA results, was particularly unexpected and has significant implications for future validation work and, perhaps, teacher education. Specifically, there is a significant amount of literature describing the lack of fraction CK of PSTs (Depaepe et al., 2015; Tröbst et al., 2018). This literature base led us to believe that a sample of junior and senior PSTs would provide a wider range of demonstrated PCK than what we observed. Instead, only 23.5% of participants demonstrated "lower than average" PCK for fractions.

Implications for Mathematics Teacher Educators

This article presents validity evidence from the pilot of our PCK-Fractions measure. Despite these results serving as the first in a series of validity argument studies for this measure, there are clear implications from the present article toward mathematics teacher education. First and foremost, the revised construct map (Table 4) provides a useful tool for teacher educators in scaffolding PSTs' experiences for engaging in such pedagogy. Such efforts can also contribute to scholarship in understanding how PSTs develop PCK for fractions, and how such development may be facilitated.

The PCK skew in scores observed in this study also has potential implications for mathematics teacher education. We caution the reader by noting these results need to be confirmed and expanded upon. It is possible that teachers may have significantly more professional knowledge than is acknowledged. It is also possible that the current measure does not include PCK items with high enough difficulty – though evidence presented in this article suggests a more pressing need for simpler tasks. As the PCK-Fractions measure, and measures like it, become available for such evaluation, it would be worthwhile to investigate which teacher education practices and professional experiences, across various teacher education programs and school districts, benefit teachers the most.

Conclusion

The purpose of this study was to construct an initial validity argument for an assessment of PSTs' PCK. Following recommendations of others (Herbst & Kosko, 2014; Copur-Gencturk et al., 2019; Hill et al., 2008), we found that designing items with a focus on student conceptions to be a beneficial approach. The validity evidence collected from our pilot study provides support for an initial validity argument. Although preliminary, evidence presented here provides a useful baseline for an initial validity argument and an improved construct map for conceptualizing PCK for fractions and studying this construct in the future.

Acknowledgments

Research reported here received support from the National Science Foundation through DRK-12 Grant #1908159. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Research reported here received support from the National Science Foundation through DRK-12 Grant #1908159.

References

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

Herbst, P., & Kosko, K. (2014). Mathematical knowledge for teaching and its specificity to high school geometryinstruction. In Research trends in mathematics teacher education (pp. 23-45) Springer, Cham.

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. https://doi.org/10.1177/0022487108324554

Bond, T. G., & Fox, C. M (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.) Routledge

Bond, T. G., Yan, Z., & Heene, M.(2020)Applying the Rasch model fundamental measurement in the human sciences (Fourth edition)Routledge.

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

Buforn, À., Llinares, S., Fernández, C., Coles, A., & Brown, L. (2020). Pre-service teachers' knowledge of the unitizing process in recognizing students' reasoning to propose teaching decisions. *International Journal of Mathematical Education in Science and Technology*, 1–19. https://doi.org/10.1080/0020739X.2020.1777333

Copur-Gencturk, Y., Tolar, T., Jacobson, E., & Fan, W. (2019). An empirical study of the dimensionality of the mathematical knowledge for teaching construct *Journal of Teacher Education*, *70*(5), 485–497. https://doi.org/10.1177/0022487118761860

Crocker, L., & Algina, J. (2006) *Introduction to classical and modern test theory*. Thomson Wadsworth

Depaepe, F., Torbeyns, J., Vermeersch, N., Janssens, D., Janssen, R., Kelchtermans, G., & Van Dooren, W. (2015). Teachers' content and pedagogical content knowledge on rational numbers: A comparison of prospective elementary and lower secondary school teachers. *Teaching and Teacher Education*, *47*, 82–92. https://doi.org/10.1016/j.tate.2014.12.009

Duncan, P. W., Bode, R. K., Lai, S., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950–963. https://doi.org/10.1016/S0003-9993(03)00035-2

Empson, S. B. (1995). Using sharing situations to help children learn fractions. *Teaching Children Mathematics*, *2*(2), 110–115. https://doi.org/10.5951/TCM.2.2.0110

Empson, S. B., Jacobs, V. R., Jessup, N. A., Hewitt, A., & Krause, G. (2020). Unit fractions as superheroes for instruction. *Mathematics Teacher: Learning and Teaching PK-12*, *113*(4), 278–286. https://doi.org/10.5951/MTLT.2018.0024

Erdem, E. (2016). Prospective middle school mathematics teachers'(PMTs) content knowledge about concepts 'fraction' and 'rational numbers' *Journal of Education and Training Studies*, *4*(6), 80–92. https://doi.org/10.11114/jets.v4i6.1431

Hackenberg, A. J., Norton, A., & Wright, R. J (2016) *Developing fractions knowledge* Sage.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *41*(5), 513–545. https://doi.org/10.5951/jresematheduc.41.5.0513

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

Huang, T. W., Liu, S. T., & Lin, C. Y (2009) Preservice teachers' mathematical knowledge of fractions. *Research in Higher Education Journal*, *5*, 1.

Hunting, R. P., & Sharpley, C. F. (1988). Fraction knowledge in preschool children. *Journal for Research in Mathematics Education*, *19*(2), 175–180. https://www.jstor.org/stable/749411

Izsák, A., Jacobson, E., & Bradshaw, L. (2019). Surveying middle-grades teachers' reasoning about fraction arithmetic in terms of measured quantities. *Journal for Research in Mathematics Education*, *50*(2), 156–209. https://doi.org/10.5951/jresematheduc.50.2.0156

Izsák, A.(2008).Mathematical knowledge for teaching fraction multiplication. Cognition and instruction, 26(1), 95–143. https://doi.org/10.1080/07370000701798529 .

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M., & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151. https://doi.org/10.1080/00461520701416231

Kazemi, F., & Rafiepour, A. (2018). Developing a scale to measure content knowledge and pedagogy content knowledge of in-service elementary teachers on fractions. *International Journal of Science and Mathematics Education*, *16*(4), 737–757. https://doi.org/10.1007/s10763-016-9792-0

Khakasa, M. M. O. C., & Berger, M. (2016). Status of teachers' proficiency in mathematical knowledge for teaching at secondary school level in Kenya. *International Journal of Science and Mathematics Education*, *14*(2), 419–435. https://doi.org/10.1007/s10763-015-9630-9

Krupa, E. E., Bostic, J. D., & Shih, J. C. (2019). Validation in mathematics education: An introduction to quantitative measures of mathematical knowledge: Researching instruments and perspectives. In J. D. Bostic, E. E. Krpua, & J. C. Shih (Eds.), *Quantitative measures of mathematical knowledge: Researching instruments and perspectives* (pp. 1–13). Routledge.

Linacre, J. M. (2021) *A user's guide to WINSTEPS, MINISTEP Rasch-model computer programs: Program manual 4.8.0*. https://www.winsteps.com/winman/

Linacre, J. K. (1994). Sample size and item calibration stability Rasch Measurement Transactions,7(4), 328.

Post, T. R., Wachsmuth, I., Lesh, R., & Behr, M. J. (1985). Order and equivalence of rational numbers: A cognitive analysis. *Journal for Research in Mathematics Education*, *16*(1), 18–36. https://doi.org/10.2307/748970

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Sick, J. (2010). Rasch measurement in language education part 5: Assumptions and requirements of rasch measurement. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *14* (2), 23–29. https://hosted.jalt.org/test/sic_5.htm

Tröbst, S., Kleickmann, T., Heinze, A., Bernholt, A., Rink, R., & Kunter, M. (2018). Teacher knowledge experiment: Testing mechanisms underlying the formation of preservice elementary school teachers' pedagogical content knowledge concerning fractions and fractional arithmetic. *Journal of Educational Psychology*, *110*(8), 1049–1065. https://doi.org/10.1037/edu0000260

Tyminski, A. M., Simpson, A. J., Land, T. J., Drake, C., & Dede, E. (2020). Prospective elementary mathematics teachers' noticing of children's mathematics: A focus on extending moves. *Journal of Mathematics Teacher Education*. https://doi.org/10.1007/s10857-020-09472-2

Tzur, R. (1999). An integrated study of children's construction of improper fractions and the teacher's role in promoting that learning. *Journal for Research in Mathematics Education*, *30*(4), 390–416. https://doi.org/10.2307/749707

Wilson, M. (2005) *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wilson, P. H., Edgington, C. P., Nguyen, K. H., Pescosolido, R. C., & Confrey, J. (2011). Fractions: How to fair share. *Mathematics Teaching in the Middle School*, *17*(4), 230–236. https://doi.org/10.5951/mathteacmiddscho.17.4.0230

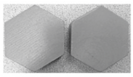
Wright, B. D. (1991). Scores, reliabilities, and assumptions. *Rasch Measurement Transactions*, *5* (3), 157–158. https://www.rasch.org/rmt/rmt53a.htm

Wright, B. D., & Stone, M. H. (2004) *Making measures*. The Phaneron Press.

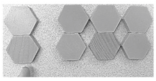
Appendix. Example items from the PCK-Fractions measure. See F01 (Table 1), F08a-b (Figure 4) & F14 (Figure 2) for additional examples

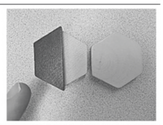
F21

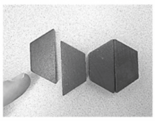
Ms. Crusher asked her students to show one-fourth of two yellow hexagons using a set of fraction pattern blocks.

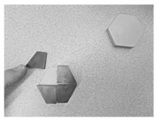


Of the responses below, which best demonstrates the incorrect conception that a whole must be one discrete object?

① 

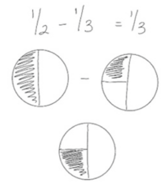
③ 

② 

④ 

F09

Ron was asked to find $\frac{1}{2}-\frac{1}{3}$, he drew the picture below and said ***"1 of 2 parts minus 1 of 3 parts is 1 of 3 parts".***



Based on his response, which of the following statements best explains his reasoning?

① Ron does not understand partitioning a whole into equal parts.

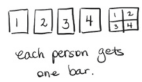
② Ron understands how to partition a whole into equal parts but does not understand what the parts mean.

③ Ron understands how to make parts but does not understand what the whole means.

④ Ron is using a fair share approach on a fraction subtraction task.

F07

John was asked to show and tell how much each person gets if four people share five chocolate bars equally. John's response is shown below:



Which statement below best describes John's reasoning?

① John correctly partitions the chocolate bar but does not provide a fraction as an answer.

② John correctly partitions the chocolate bar and provides a fraction as an answer.

③ John incorrectly partitions the chocolate bar but does not provide a fraction as an answer.

④ John incorrectly partitions the chocolate bar and provides a fraction as an answer.