Development and validation of a survey instrument for measuring pre-service teachers' pedagogical content knowledge

Martin, David and Jamieson-Proctor, Romina

Development and validation of a survey instrument for measuring pre-service teachers' pedagogical content knowledge

David Martin & Romina Jamieson-Proctor

To cite this article: David Martin & Romina Jamieson-Proctor (2019): Development and validation of a survey instrument for measuring pre-service teachers pedagogical content knowledge, International Journal of Research & Method in Education, DOI: 10.1080/1743727X.2019.1687669

In Australia, one of the key findings of the Teacher Education Ministerial Advisory Group was that not all graduating pre-service teachers possess adequate pedagogical content knowledge (PCK) to teach effectively. The concern is that higher education providers working with pre-service teachers are using pedagogical practices and assessments which are not informed by research. Due to its multifaceted nature, PCK is difficult to measure. In the case of mathematics PCK, some researchers have used multiple-choice questions while others use constructed-response formats depending on the specific attribute of PCK being measured. In either case, researchers need to use instruments that are appropriate to each task. This paper reports on the construction and reliability of a survey instrument used in a larger study to measure pre-service teachers' mathematics PCK. Using four stages of survey development, in addition to inter-rater agreement measures and Rasch modelling techniques, data analyses found high levels of validity and reliability for the Mathematics Pedagogical Content Knowledge Instrument (MPCKI).

Keywords: pre-service teachers; mathematics pedagogical content knowledge; construct validity; Rasch modelling

Introduction

It is suggested from the literature (Carroll and Foster 2010; National Research Council 2010; Office for Standards in Education 2005; Teacher Education Ministerial Advisory Group 2014) that knowledge of subject matter and pedagogical approaches are important elements of teacher effectiveness and therefore are a focus of pre-service teacher education programs. Shulman (1987, 8) defines "the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organised, represented… and presented for instruction" as pedagogical content knowledge, or PCK. The assessment of pre-service teachers' PCK provides data that teacher educators can use in order to report on the pre-service teachers' learning (Tatto et al. 2008), as well as plan future learning opportunities that cater for their often diverse needs. In order to effectively measure PCK, or any construct, it is important to first conceptualize the construct (Johnson and Christensen 2017). Additionally, survey instruments need to be appropriate to each task since the conclusions which are formed from the analysis of the survey results will only be as good as the validity and reliability of the items comprising the instrument (Bond and Fox 2007).

The items created for the Mathematics Pedagogical Content Knowledge Instrument (MPCKI), were developed for use in a study (Martin 2017) which measured the impact of a problem-based learning (PBL) approach on pre-service teachers' mathematics PCK. The items were chosen for their alignment with the content of the mathematics subject in which the participants, all third-year pre-service teachers, were enrolled. The semester-long mathematics curriculum and pedagogy subject aimed to provide the pre-service teachers with the knowledge and skills to effectively teach the Australian Mathematics Curriculum areas of Algebra, Measurement, Geometry and Probability and Statistics at the primary school level.

The focus of the subject was the development of pedagogical strategies based on a carefully constructed and well-researched theoretical pedagogical model. However, concurrently, it has been noted that students generally also develop their mathematics content knowledge as they explore effective ways to teach mathematics and reflect on their own mathematics education journeys.

The stages used to design the MPCKI, align with Senocak's (2009) framework for the development of the Problem-based Learning Environment Inventory (PBLEI). Senocak's framework involves four distinct stages: item formation, content validation, construct validation, and reliability calculation. In addition to testing for validity and reliability, inter-rater agreement was also measured as a fifth stage, so that cross checks and a consensus of the best fitting answers to the MPCKI questions could be established. Inter-rater agreement is the percentage of how frequently two or more evaluators agree on the same rating to an identical situation (Gisev, Bell, and Chen 2013; Graham, Milanowski, and Miller 2012).

Stage 1: Item formulation

Item formation requires an extensive review of the literature (Senocak 2009) so that researchers can have confidence that the developed instrucment is constructed in a way that ensures it measures what they are intending to measure (validity), and that they do this consistently (reliability) (Pedersen, Hill, and Callingham 2015).

Conceptualisation of pedagogical content knowledge (PCK)

The degree of mastery of subject matter, or content knowledge (CK), and how to deliver that content using pedagogical knowledge (PK) are acknowledged as major factors of a teacher's effectiveness. In terms of distinction, CK is the knowledge *of* the subject matter and also *about* the subject matter within the context of teaching that content (Ball 1988). PK is characterised as a significant component of teaching practice and can be further subdivided into *instructional planning*, *student learning* and *curricular knowledge* (Schmidt et al. 2007).

Shulman (1986) conceptualised PCK as a teacher's special form of professional understanding used to formulate the content knowledge and his or her concepts into the most powerful representations, analogies, illustrations, explanations and demonstrations which make them comprehensible to students. A year later, Shulman defined PCK as "the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organised, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (Shulman 1987, 8).

Researchers that followed, such as Cochran, King, and DeRuiter (1993); Gess-Newsome (1999); Grossman (1990); Marks (1990); (Ball 2000; Ball, Lubienski, and Mewborn 2001) and Chick (2003) are credited with progressing the conceptualization of PCK as the ability of teachers to make the content instructional (Segall 2004). Segall also gave credit for their work in determining the relationship between how teachers integrate content and pedagogy in classrooms. Mirroring many prior suppositions regarding PCK, Segall also posits the question, can CK and PK be separated, or are they inherently joined? In response, Segall proposes PCK should be regarded as more than a matter of teachers knowing their subject matter and using that knowledge to summon the best pedagogy to make the content knowledge instructional.

Chick (2003) articulated PCK (of mathematics specifically) as a form of teacher knowledge which goes beyond mastery of the subject to incorporate how the mathematics CK is used in teaching. "This includes a knowledge of the representations and models that can be used to depict important mathematical principles, understanding the connections between fundamental mathematical ideas and knowledge of significant examples and counterexamples” (2003, 1).

Chick, Pham, and Baker (2006) conceptualised a framework for analysing mathematics PCK which is grouped into three categories with mathematics teaching as the context. In this framework, PCK is considered multifaceted with a mix of content and pedagogy regarded as: (a) completely intertwined with pedagogy and content, (b) content knowledge in a pedagogical context, and (c) pedagogical knowledge within a content context. The first category of the framework, PCK, is consistent with other conceptualizations of PCK.

Ball, Thames, and Phelps suggest that few studies tested whether there are indeed “distinct bodies of identifiable content knowledge that matter for teaching” (2008, 4). Their research group chose to investigate specifically what teachers need to know about the mathematics content they are asked to teach; and, how and where might they use such mathematical knowledge in practice. Based on those two questions and their analyses of videos of teaching practice and subsequent design measures of mathematical knowledge formulated from those videos, they developed a working definition of ‘mathematical knowledge for teaching’. By this phrase they meant, “the mathematical knowledge that teachers need to carry out their work as teachers of mathematics” (2008, 4). Further, they proposed a modification to Shulman’s initial categories: subject matter knowledge and pedagogical content knowledge. Based on their analysis they hypothesised that Shulman’s (1986) two categories, subject matter knowledge and pedagogical content knowledge can be further subdivided. Subject matter knowledge they concluded was comprised of *common content knowledge* and *specialised content knowledge,* while pedagogical knowledge according to Ball et al. (2008) requires knowledge of content, curriculum and teaching.

Instruments used for measuring pre-service teachers’ mathematics PCK for teaching have been identified in the literature (Callingham and Beswick 2011; Cheang et al. 2007; Tatto et al. 2008). Prior to each teams’ development of their respective questionnaires, each research team first set out to define the multi-faceted construct of mathematics PCK.

For example, the team of Cheang, Yeo, Chan and Lim-Teo (2007) agreed the construct of mathematics PCK included:

(1) a teacher’s own understanding of mathematical structure and connections,
(2) a teacher’s knowledge of a range of alternative representations of concepts for purpose of explanation,
(3) a teacher’s ability to analyse the cognitive demands of mathematical tasks on learners, and
(4) a teacher’s ability to understand and take appropriate action for children’s learning difficulties and misconceptions.

The researchers of the CEMENT team (Callingham and Beswick 2011) formulated a framework of mathematics PCK grounded on pre-service teachers being able to:

(1) identify students’ errors or misconceptions,
(2) construct or use tasks and tools for developing students’ understanding,
(3) know and apply of a range of representations of a particular mathematical idea, and
(4) explain ideas to students.

The Teacher Education and Development Study in Mathematics (TEDS-M) project (Tatto et al. 2008) was the first cross-national study to bring together international experts in mathematics education, research, curriculum, instruction and assessment to develop assessment items for measuring pre-service teachers' mathematics PCK. One of the first challenges facing the TEDS-M project developers was defining and operationalising the domain of PCK in an international context. To tackle this challenge expert representatives from the participating countries met to develop and approve a collective conceptualization of PCK. The result produced two internationally accepted sub-domains of PCK:

(1) curricular knowledge and knowledge of planning for mathematics teaching and learning, and
(2) knowledge of enacting mathematics for teaching and learning (Tatto et al. 2008).

Considering the definitions of PCK from the studies (Callingham and Beswick 2011; Cheang et al. 2007; Tatto et al. 2008), it was observed that once you take the references to any components of CK and PK from the research teams' definitions of PCK, what generally remains is the teacher's ability to understand how students think and learn, and then how to act on that knowledge. In essence, it is the ability of a teacher to:

(1) know and recognise the cognitive demands of tasks on students,
(2) identify any learning difficulties and misconceptions students might have, and
(3) be able to appropriately construct and perform tasks which develop understanding and address students' misconceptions and learning difficulties.

Therefore, using mathematics as the discipline, PCK is defined for the purposes of the current study as an integration of 3 elements:

(1) content knowledge,
(2) pedagogical knowledge and
(3) applying planning, teaching and assessment for learning based on knowledge of students' cognitive abilities, understanding and potential misunderstandings.

In addition to measuring the pre-service teachers' mathematics PCK using the MPCKI, the larger study in which the MPCKI was developed, and used, employed eight PCK exam questions to measure enacting of mathematics PCK (element 3 above). Hence, the MPCKI in conjunction with the end-of-semester PCK exam questions primarily measured mathematics PCK in two forms. The MPCKI used multiple choice questions to measure how a teacher would a) approach a particular teaching strategy, b) respond to students' ways of thinking, and c) address typical student misconceptions. The end-of-semester exam used extended response PCK questions that required the pre-service teachers to design activities which addressed applying planning, representation of ideas, assessment for learning and difficulties children experience in particular mathematics concepts and skills.

Having identified the conceptual framework for PCK, the following sections describe the process of item formation of the multiple-choice survey questions for the MPCKI and tests of validity and reliability involved in confirming the instrument was fit for purpose.

Item formation of the Mathematics Pedagogical Content Knowledge Instrument (MPCKI)

Following the extensive research of mathematics PCK, three studies formed the basis for the development of the MPCKI items with contributions and permission from Callingham and

Beswick (2011) and Cheang et al. (2007), and contributions and acknowledgement of the use of several released items of the TEDS-M (Australian Council for Educational Research for the TEDS-M International Study Centre 2011).

Cheang et al. (2007), from the National Institute of Education in Singapore, initiated a project titled *Knowledge for Teaching Primary Mathematics* (MPCK Project). For the project, they developed a 16-item, short answer questionnaire to measure specific aspects of mathematics PCK. Eight of the 16 items covered topics associated with whole numbers, fractions or decimals, which were not the content covered in the specific mathematics subject the pre-service teachers were enrolled in. Therefore, these eight items were not considered for use in the MPCKI. The other eight short answer items were a mix of measurement and geometry questions that aligned with the content of the subject. Of those eight items, several were noted by the researcher as potential items for inclusion in the MPCKI due to their alignment with the learning outcomes of the mathematics curriculum and pedagogy associated with the study (Martin 2017).

The CEMENT instrument (Callingham and Beswick 2011) consists of three scales: teachers' beliefs about mathematics and its teaching, mathematics content knowledge and mathematics PCK. Of the three scales, one was relevant to this study. The CEMENT team's mathematics PCK scale, which used a multiple-choice format, included item topics associated with measurement, geometry and probability, as well as whole numbers, fractions and decimals. Therefore, not all items on the instrument were deemed suitable for use. However, a select number of measurement, geometry and probability multiple-choice questions were considered as potential items for inclusion in the MPCKI.

The third useful instrument for measuring future teachers' mathematics PCK came from the TEDS-M Study. In 2011, a set of items from the TEDS-M test were released. The test bank is comprised of 34 questions, 10 of which are mathematics PCK questions, which required either a constructed or multiple-choice response. Of the 10 mathematics PCK questions, five were perceived by the researcher as potential survey items for the MPCKI due to their alignment with the content and learning objectives of the mathematics subject.

Items used from the three contributing instruments varied in number and type. Specifically, three of the short answer items from Cheang et al. (2007) MPCK Project were modified into classroom teaching scenarios requiring responses to four or five multiple choice questions. For example, the stem for question 14 of Cheang et al. (2007) MPCK Project questionnaire asks,

> "When teaching children about measurement for the first time, Mdm Ho prefers to begin by having the children measure the width of their book using paper clips, and then again using pencils. Why do you think she prefers to do this rather than simply teaching the children how to use a ruler?" (2007, 53).

Similarly, this stem was used as one of the items on the 2008 TEDS-M and released in 2011. The item and included indicative responses consisted of four categories - correct response, partially correct response, incorrect response, and non-response (Australian Council for Educational Research for the TEDS-M International Study Centre 2011). This stem and related indicative sample responses from the TEDS-M formed the basis of the four multiple choice items for this MPCKI question. The scenario was restructured on the MPCKI so that the pre-service teachers needed to respond to four multiple-choice questions in relation to the scenario (Figure 1). Another set of MPCKI items situated in a different scenario can be found in Appendix A.

In total, the MPCKI is designed with 54 multiple-choice questions identified as measuring pre-service teachers' mathematics PCK. Inclusive of the multiple-choice questions shown in Figure 1 and Appendix A, there are 54 questions in the MPCKI which are situated in 12 scenarios of classroom teaching (three algebra, three measurement, three geometry, and three statistics & probability).

When teaching children length measurement for the first time, Mrs. Brown prefers to begin by having the children measure the width of their book using paper clips, then again using pencils.

Below, there are four possible reasons why Mrs. Brown would use this strategy to teach length measurement. For each, indicate whether you believe it is a *Correct reason, Partially correct reason,* or *Incorrect reason*.

	Correct reason	**Partially correct reason**	**Incorrect reason**
Using familiar/different units enables understanding of what measurement is and that any object/unit with length can be used to measure.	[x]	[ ]	[ ]
Using non-standard units of length to measure gives differing numbers of units for the same length and shows that we need standard units.	[ ]	[x]	[ ]
The teacher knows that the students will enjoy their work if they can use hands-on materials.	[ ]	[ ]	[x]
Using objects of different lengths helps children learn how to decide which unit/object is the most appropriate to measure a given length.	[x]	[ ]	[ ]

Figure 1. Example Measurement Items in the MPCKI.

Each teaching scenario contains four to five of the 54 multiple-choice questions, with each item consisting of three response choices. The response choices were spread across three categories for each scenario. The correctness of each response choice was based on expert understanding and research into the most appropriate way to teach each mathematical concept described in the scenario.

Table 1 provides the list of the 12 scenarios of classroom teaching problems used in the MPCKI and identifies the contributing instruments for each, with one algebra scenario designed by the researcher.

Table 1. Items in the MPCKI, Identifying the Original Instruments they were drawn from.

	MPCKI Item	MPCK Project (Cheang et al., 2007)	CEMEMT (Callingham & Beswick, 2011)	TEDS-M (Tatto et al., 2008)	Designed by researcher
1	Algebra				√
2	Geometry	√			
3	Measurement		√		
4	Probability & Statistics			√	
5	Geometry		√		
6	Geometry		√		
7	Probability & Statistics		√		
8	Algebra			√	
9	Algebra			√	
10	Measurement			√	
11	Measurement	√			
12	Probability & Statistics			√	

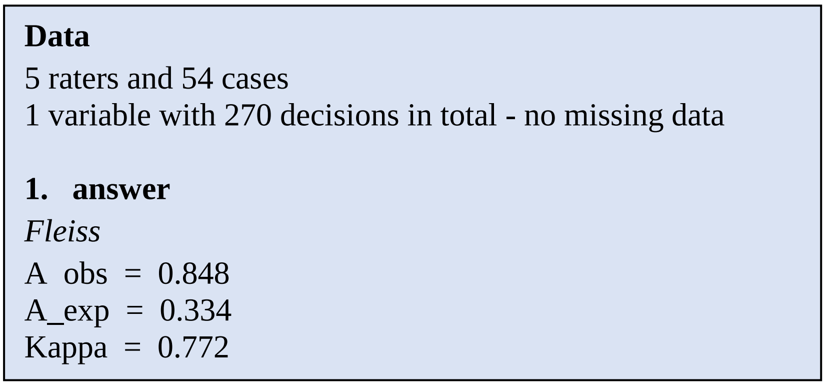
Following the selection and modifications of the MPCKI's 12 mathematics teaching scenarios, the 54 individual items were tested for content validity, construct validity and reliability with the pre-service teachers (N=238).

Stage 2: Content validation

Content validity is a subjective estimate on the degree to which one might view how well a measure operationalises a construct (Drost 2011). To test the items for content validity, the MPCKI was examined by a sample of five experts in mathematics PCK. All five were mathematics education academics from three Australian universities who were chosen on the basis of their experience in the field of mathematics PCK for pre-service teachers. The panel of experts were asked to provide critical feedback with suggestions for improvement. The objective was to have each expert provide their respective judgments regarding whether they felt the MPCKI appeared to be a good measure of mathematics PCK (content validity) specifically for pre-service teachers. After an examination of their feedback and changes made to the survey, each of the experts agreed the 54 items within the MPCKI measured mathematics PCK.

Stage 3: Inter-rater agreement

The process for measuring inter-rater agreement requires that each mathematics PCK expert complete the MPCKI (International Test Commission 2014). There are three common methods for computing inter-rater agreement: intra-class correlation, the percentage agreement measure, and the Cohen's kappa measure (Graham et al. 2012; Stemler 2004). The intra-class index is useful when there are five or more rating categories. Since the items in the MPCKI are not composed of five or more categories, the intra-class correlation index was deemed unsuitable for the inter-rater agreement test. The percentage agreement measure is not a statistical analysis which eliminates correlation coefficients between raters who agree by chance (International Test Commission 2014). Thus, the index was considered an unacceptable measure for determining the inter-rater agreement amongst the five experts' answers to the 54 questions of the MPCKI. Cohen's kappa measures the inter-rater agreement between two raters using a two-level rating scale of nominal/categorical variable, and where agreement due to chance is factored out (Graham et al. 2012). Due to the restriction of measuring between only two raters, the Cohen's kappa test was unsuitable for this study. Fortunately, Cohen's kappa has been extended so that the number of raters can be more than two. Fleiss' kappa can be used when nominal categories are assessed by multiple raters and, like Cohen's kappa, corrects for agreement being reached by chance (Gisev et al. 2013). Fleiss' kappa, however, does not assume that the raters have all assessed all items (2013). This was not a concern for this study as each of the five experts agreed to assess all 54 items. As a result, it was determined that the agreement percentage among the mathematics PCK experts' ratings would be best determined using Fleiss' kappa. Fleiss' kappa values were calculated to determine inter-rater agreement with multiple raters using an online calculator (https://mlnl.net/jg/software/ira/) (Geertzen 2012). Discussions with each rater were used to come to agreement regarding the appropriateness of responses for each item. The Fleiss kappa value was then calculated to be 0.77 (Figure 2).



Data
5 raters and 54 cases
1 variable with 270 decisions in total - no missing data

1. answer
Fleiss
A obs = 0.848
A_exp = 0.334
Kappa = 0.772

Figure 2. Fleiss' Kappa Value for the Inter-rater Agreement of the 54 MPCKI Items.

This level of agreement is considered to be excellent and beyond chance:

> For most purposes, [Kappa] values greater than 0.75 or so may be taken to represent excellent agreement beyond chance, values below 0.40 or so may be taken to represent poor agreement beyond chance, and values between 0.40 and 0.75 may be taken to represent fair to good agreement beyond chance. (Fleiss, Levin, and Paik 2003, 604)

Stage 4: Construct Validity

One approach to establish construct validity is to examine the fit of the instruments' items to the underlying construct using Rasch modelling techniques (Bond and Fox 2007). Fit indices returned from Rasch analyses provide indicators of unidimensionality and evidence of items within a test which do not fit the representation of the construct.

To establish construct validity of the MPCKI, all 54 items were loaded into Qualtrics electronic survey software and presented to 238 pre-service teachers. For studies involving the Rasch model, a minimum of 20 items and a sample size of 200 examinees are sufficient (Wright and Stone 1979; Zubairi and Kassim 2006). Item scoring was dichotomous (right or wrong) and the Qualtrics software automatically scored the items. Specifically, each item had three response choices (see Figure 1 and Appendix A). The correct response was coded as 1. The partially correct and incorrect responses were coded as 0. Rasch model analyses of the data set were conducted using WINSTEPS Version 3.81.0 (Linacre 2014). The initial analysis was run to test for unidimensionality and to obtain the fit levels of all the MPCKI test items against the latent trait. A latent trait is the fictitious straight line used to represent the theoretically perfect representation of the construct (Bond and Fox 2007). Along with providing this evidence as a data set, the Rasch model can illustrate the results graphically. Figure 3 illustrates the fit statistics in the form of the bubble chart produced by WINSTEPS.

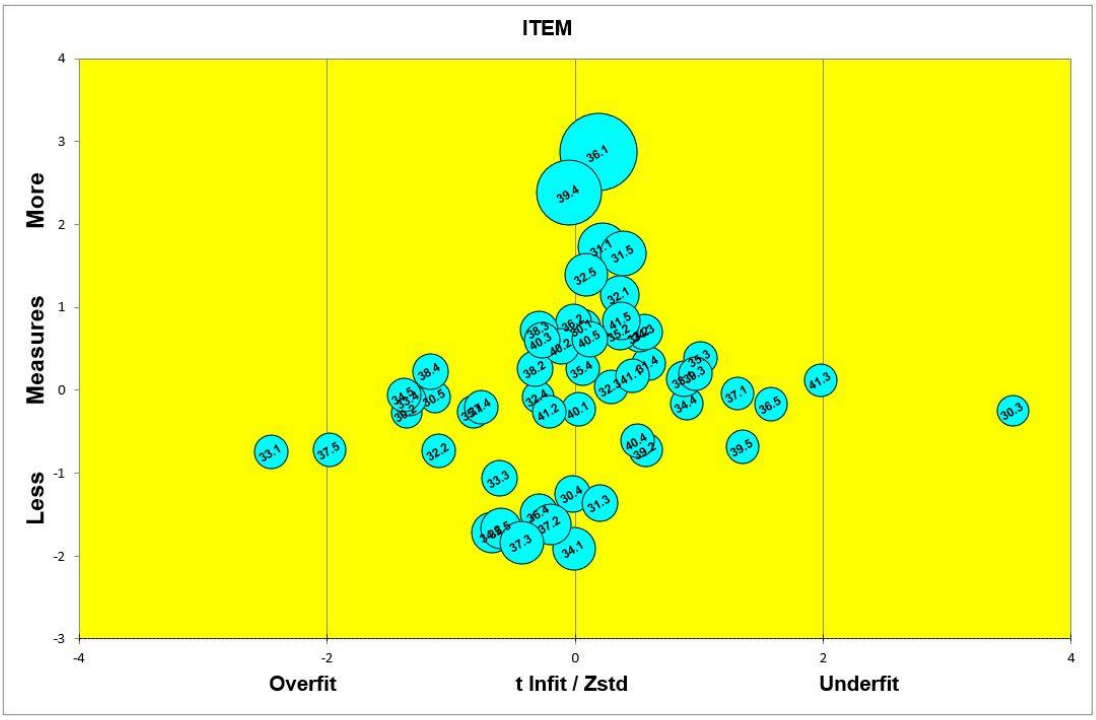


Figure 3. Bubble Chart how each of the 54 Items Fit the Scale Measured by the MPCKI.

The size of each bubble is related to the estimate of the measurement error of that item, where larger bubbles indicate greater error (standard error) of that estimate (Bond 2003). The fit of each item to the model is determined along the horizontal axis. Fit values are reported as a standardised *t* scale and acceptable values fall between -2.0 and +2.0 (Bond and Fox 2007). Fit statistics also provide indications of item misfit. Based on guidelines suggested by Bond and Fox (2007) and Linacre (2012), for this type of multiple-choice test, items with MNSQ values falling between 0.7 and 1.3 are considered to fit the model.

The item fit statistics underpinning the bubble chart reveal all but 2 items (33.1 and 30.3) fall on or within + or – 2 standard deviations and thus necessitate further investigation. The item fit statistics identified item 33.1 as overfitting and having an infit MNSQ of 0.91 and a ZSTD of -2.4. The outfit ZSTD was -2.7 and a MNSQ of .89. Hence, the overfitting item is not a threat to the validity of the scale because it is measuring the same construct and the MNSQ values fall within the 0.7 – 1.3 guideline, indicating a fit to the model (Linacre, 2012).

Items that misfit on the right-hand scale are the ones that are most random. If there are many items which misfit on the right-hand side of the scale, there is a threat to validity because it suggests that the scale may not be measuring a single construct. However, item 30.3 (see Appendix A) is the only item of concern with a ZSTD over 2.0 suggesting that with this cohort, the pre-service teachers with high PCK are rating item 30.3 hard, or the other way around; hence, there are inconsistencies in with the way the respondents are answering the question. From the point of view of the validity of the scale this may be of concern. The item statistics identified item 30.3 as having an Infit MNSQ of 1.08 and ZSTD of 3.5 and a MNSQ of 1.09 for the Outfit with a ZSTD of 3.6. Therefore, although the ZSTDs are outside of 2.0

and -2.0, the MNSQs fall comfortably within the expected range 0.7 to 1.3. If MNSQs are acceptable, the ZSTD statistic can be ignored because mean squares near 1 indicate little distortion of the measurement system, regardless of the ZSTD value (Linacre 2012). The summary fit statistics for the MPCKI scale are shown in Table 2.

Table 2. Item (N=54) Summary Fit Statistics and Separation and Reliability Indices.

	INFIT		OUTFIT	
	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1.00	.0	1.00	.0
S.D.	.04	1.0	.07	1.1
MAX	1.08	3.5	1.31	3.6
MIN	.91	-2.4	.86	-2.7
SEPARATION = 6.21;	ITEM RELIABILITY = .97			

The MNSQ infit and outfit statistics for all 54 items of the MPCKI fell in the 0.9 - 1.1 and 0.9 - 1.3 range, respectively; that is, these items appear to be behaving as expected and each item is a productive measure of the construct in the sense they are contributing information to the scale.

The mean square infit and outfit statistic for the 54 items collectively were 1.00 and 1.00; therefore, the expected mean square value of 1.00 was achieved. The ZSTD for the infit and outfit are the standardization of the fit scores, reported in various interval-scale forms such as *t* or *z*. Basically, the mean squares are transformed so they are distributed like *t*, with an expected mean of 0 (Bond and Fox 2007), which again was achieved. As a result, the overall fit of the test to the model indicate the 54 items were a satisfactory representation of the construct mathematics PCK as determined by the study.

Stage 5: Reliabilty Calculation

The item separation and item reliability values returned from the Rasch analysis are provided in Table 2. Item separation is used as an indication of the item hierarchy and how well the items are separated by the participants taking the test (Linacre 2012; Wright and Stone 1999). Good item separation (> 3) and good item reliability (> 0.90) indicate the sample is large enough to draw conclusions about the hierarchy of item difficulties, and that the person sample size is large enough (Linacre 2012). The item separation (6.21) and item reliability (0.97) values signify that the order of item estimates defines a distinct hierarchy along the dimension and can be relied upon to be replicated with other appropriate samples of pre-service teachers (Bond and Fox 2007).

In terms of item difficulty, the bubble chart (Figure 3), represents the order of item difficulty vertically with the upper most items representing the more difficult questions for the 238 pre-service teachers. Conversely, the items nearest to the bottom of the chart represent the less difficult items for the 238 pre-service teachers. The scaled intervals along the vertical axis illustrate the difficulty level of the items in relation to the mean in terms of logit values. “Based on the logic of order, the Rasch analysis software program performs a logarithmic transformation on the item and person data to convert the ordinal data to yield interval data” (Bond and Fox 2007, 29). Specific to this analysis, the actual performance of the 238 pre-service teachers on the 54 MPCKI items determined the interval sizes which allowed for the difficulty of the items to be placed along an interval scale. Furthermore, measured estimates of the 54 items provide clarification of how far apart these items are, relative to each other.

Not only can the Rasch model provide evidence of a single dimension along with the difficulty order of the items along that dimension in a bubble chart, but WINSTEPS can also produce a variable map which illustrates the measures by which the items varied in difficulty. In this way the performance of the pre-service teachers tested can be demonstrated on the same latent trait as the 54 MPCKI items, “and with similar meaningful ability distances revealed between the students” (Bond and Fox 2007, xiii). The person-item mapping in Figure 4 illustrates the difficulty level of the items in relation to the mean (+M, set at 0) in terms of logit values, alongside the pre-service teachers’ ability levels with their mean denoted at M.

```
MEASURE                 PERSON   Map    ITEM
                      more        |       rare
   3                              +
                                  |
                                  |     36.1
                                  |
                                  |
                                  |
                                  |
                                  |     39.4
                                  |
                                  |
                                  |
                                  |
   2                              +T
                                  |
                                  |
                                  |     31.1
                                  |     31.5
                                  |
                                  |
                                  |     32.5
                                  |
                                  |
                                  |     32.1
                                  |
   1                              +S
                                  |
                                  |     36.2    41.5
                                  |     30.1    38.3
                                  |     33.2    34.3    35.2
                                  |     40.3    40.5
                                  |     40.2
                        ##    T   |     35.3
                         #        |     31.4
                         #        |     35.4    38.2    38.4
                       ###        |     38.1    39.3    41.1
                    ######    S   |     32.3    41.3
   0                ######        +M    37.1
                    ######        |     30.5    32.4    33.4    34.5
                ##########        |     34.4    36.5    37.4
                 #########    M   |     30.2    30.3    35.1    40.1    41.2
                    ######        |
                         #        |
                     #####        |
                       ###    S   |     40.4
                       ###        |     39.5
                       ###        |     32.2    33.1    37.5    39.2
                        ##        |
                         #    T   |
  -1                     #        +S
                                  |     33.3
                                  |
                                  |     30.4
                                  |     31.3
                                  |
                                  |     36.4
                                  |     37.2
                                  |     35.5
                                  |     34.2
                                  |     37.3
                                  |     34.1
  -2                              +T
                           less   |   frequent
```

Figure 4. Person-Item Map for the MPCKI Data (N=238).

The MPCKI items are to the right of the vertical line, and pre-service teachers are to the left denoted by a #, where each # represents three pre-service teachers. According to the item statistics returned from the Rasch analyses, the 54 items were ordered by average ability (measure) that ranged from -1.91 for item 34.1 to 2.87 for item 36.1. That is, in terms of level of difficulty the 54 items were distributed across most of the -2 to +3 logit range suggesting good separation.

In relation to the participant sample (of pre-service teachers), given the range of varying difficulties of the items, the person mapping in Figure 4 shows lack of good separation suggesting the sample of pre-service teachers is a relatively homogenous group in terms of mathematics PCK knowledge. An examination of the statistics (Table 3) underpinning the person mapping provided by Rasch analyses shows the mean square infit and outfit values of 1.00 and 1.00, respectively, suggesting the pre-service teachers are behaving as expected by the Rasch model.

Table 3. Participant (N=231) Separation and Reliability Indices.

	INFIT		OUTFIT	
	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1.00	-0.1	1.00	.0
S.D.	.15	1.2	.24	1.2
SEPARATION = .42;	ITEM RELIABILITY = .15			
LACKING RESPONSES	7 person			
VALID RESPONSES	94.6% (APPROXIMATE)			
CRONBACH α PERSON Raw Score “TEST” RELAIBILITY = .65 (approximate)				

In essence, the items are well separated across six ability levels, but the items are not discriminating well between the high and low achievers. However, a person reliability Cronbach alpha value of 0.65, from a Rasch perspective, suggests there is correlation among the ways that these pre-service teachers have answered the 54 items. At the pre-test this type of person separation, and a lower person ability mean (M) compared to the item difficulty mean (+M), is what you might look for in a pre-test post-test situation. For example, in terms of this study, it is hypothesised that after learning has taken place and been measured, a greater separation between the control group and the PBL treatment groups’ level of mathematics PCK should occur.

Conclusions and implications

This paper discussed the development and validity testing of the 54-item survey instrument, MPCKI, with pre-service teachers (N=238), designed to measure pre-service teachers’ mathematics PCK. Rasch analyses returned statistics which suggest the 54 items of the MPCKI can be relied upon to measure mathematics PCK, as described in the study (Martin 2017) and suggests the order of item estimates could be relied upon to be replicated with suitable participants (Bond and Fox 2007).

Limitations and suggestions for future research

Limitations of the study (Martin 2017) associated with using the MPCKI in a pre- post-test design is testing and instrumentation threats. A testing threat suggests that an improvement on a post-test is a result of having taken a matched pre-test. Conversely, instrumentation threat can occur in quantitative measurement when the pre-test instrument and post-test instrument are not equivalent in terms of measuring the same construct (Fraenkel, Wallen, and Hyun 2015; Johnson and Christensen 2017; Withrow 2013).

Further Rasch analyses may provide a means to reduce these threats. In addition to Rasch analyses providing indicators of unidimensionality and evidence of items within a test which do not fit the representation of the construct, WINSTEPS can show the order of the items' difficulty along the latent trait, as well as the amount by which the items vary in difficulty (Bond and Fox 2007). These related item statistics revealing the amount by which the items vary in difficulty have practical applications; allowing the items to be separated, based on their levels of difficulty to create a pre-test MPCKI and post-test MPCKI for the study. This is because items which band together on the same logit levels suggest these items are measuring approximately the same particular intellectual skill. Although validating the MPCKI into a pre-test and post-test is not the focus of this paper, it should be acknowledged that this feature of Rasch allows for the possible development of two variations of the MPCKI which can then be validated as measuring the same constructs at the same or similar difficulty level (Hill and Ball 2004). Hence further research to reduce the potential testing and instrumentation threats is warranted.

References

Australian Council for Educational Research for the TEDS-M International Study Centre. 2011. "*Released Items: Future Teacher Mathematics Content Knowledge (MCK) and Mathematics Pedagogical Content Knowledge (MPCK) - Primary*." International Association for the Evaluation of Educational Achievement. ERIC

Ball, D. 1988. "Knowledge and Reasoning in Mathematical Pedagogy: Examining what Prospective Teachers bring to Teacher Education." PhD dissertation. Michigan State University. https://deborahloewenbergball.com/publications#Papers

Ball, D. 2000. "Bridging Practices: Intertwining Content and Pedagogy in Teaching and Learning to Teach." *Journal of Teacher Education* 51: 241-247. SAGE

Ball, D., S. Lubienski and D. Mewborn. 2001. "Research on Teaching Mathematics: The Unsolved Problem of Teachers' Mathematics Knowledge." In *Handbook of Research on Teaching*, edited by V. Richardson, 433-456. New York: Macmillan. Google Scholar

Ball, D., M. Thames, and G. Phelps. 2008. "Content Knowledge for Teaching: What Makes it Special?" *Journal of Teacher Education* 59 (5): 389-407. SAGE

Bond, T. 2003. "Validity and Assessment: a Rasch Measurement Perspective." *Metodologia de las Ciencias del Comportamiento* 5 (2): 179-194. Google Scholar

Bond, T., and C. Fox. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum. Google Scholar

Callingham, R., and K. Beswick. 2011. "Measuring Pre-service Teachers Knowledge of Mathematics for Teaching." Paper presented at the annual conference for the Australian Association for Research in Education, Hobart, Tasmania. AARE

Carroll, T., and E. Foster. 2010. *Who Will teach? Experience Matters*. Washington, DC: National Commission on Teaching and America's Future. ERIC

Cheang, W. K., J. Yeo, E. Chan, and S. K. Lim-Teo. 2007. "Development of Mathematics Pedagogical Content Knowledge in Student Teachers." *The Mathematics Educator* 10 (2): 27-54. NIE

Chick, H. 2003. "Pre-service Teachers' Explanations of Two Mathematical Concepts" Paper presented at the annual conference for the Australian Association for Research in Education, Auckland, NZ. AARE

Chick, H., T. Pham, and M. Baker. 2006. “Probing Teachers' Pedagogical Content Knowledge: Lessons from the Case of the Subtraction Algorithm.” Paper presented at the annual conference for the Mathematics Education Research Group of Australasia. Canberra, ACT. MERGA

Cochran, K., J. DeRuiter, and R. King. 1993. “Pedagogical Content Knowledge: An Integrative Model for Teacher Preparation.” *Journal of Teacher Education* 44 (4): 263-272. SAGE

Drost, E. 2011. “Validity and Reliability in Social Science Research.” *Education Research and Perspectives* 38 (1): 105-123. Google Scholar

Fleiss, J., B. Levin, and M. Paik. 2003. “The Measurement of Interrater Agreement.” In *Statistical Methods for Rates and Proportions,* edited by W. Shewart and S. Wilks, 598-626. 3rd ed. New Jersey: John Wiley & Sons, Inc. Wiley

Fraenkel, J., N. Wallen, and H. Hyun. 2015. *How to Design and Evaluate Research in Education.* 9th ed. New York: McGraw-Hill. Google Scholar

Gess-Newsome, J. 1999. “Pedagogical Content Knowledge: An Introduction and Orientation.” In *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education*, edited by J. Gess-Newsome and N. G. Lederman, 3-20. Boston: Kluwer Academic Publishers. Springer

Gisev, N., J. Bell, and T. Chen. 2013. “Interrater Agreement and Interrater Reliability: Key Concepts, Approaches, and Applications.” *Research in Social and Administrative Pharmacy* 9 (3): 330-338. Springer

Graham, M., A. Milanowski, and J. Miller. 2012. “Measuring and Promoting Inter-rater Agreement of Teacher and Principal Performance Ratings.” Center for Educator Compensation Reform. ERIC

Grossman, P. 1990. *The Making of a Teacher: Teacher Knowledge and Teacher Education.* Teachers College Press, Teachers College, Columbia University. SAGE

Hill, H., and D. Ball. 2004. “Learning Mathematics for Teaching: California's Mathematics Professional Development Institutes.” *Journal for Research in Mathematics Education* 35 (5): 330-351. JSTOR

International Test Commission. 2014. “ITC Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores.” *International Journal of Testing* 14 (3): 195-217. Taylor & Francis Online

Johnson, B., and L. Christensen. 2017. *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. 6th ed. Thousand Oaks: SAGE Publications.

Linacre, J. 2012. *A User's Guide to WINSTEPS: Rasch-model Computer Programs: Program Manual (Version 3.74.0)*. Citeseerx

Linacre, J. 2014. *WINSTEPS Rasch Measurement Computer Program (Version 3.81).* Chicago, IL: Winsteps.com.

Marks, R. 1990. “Pedagogical Content Knowledge: From a Mathematical Case to a Modified Conception.” *Journal of Teacher Education* 41 (3): 3-11. SAGE

Martin, D. A. 2017. The Impact of Problem-based Learning on Pre-service Teachers' Development and Application of their Mathematics Pedagogical Content Knowledge (PhD dissertation, University of Southern Queensland, Australia). https://eprints.usq.edu.au/32851/

National Research Council. 2010. *Preparing Teachers: Building Evidence for Sound Policy*. Washington, DC: Center for Education, Division of Behavioural and Social Sciences and Education. https://www.nap.edu/read/12882/chapter/1

Office for Standards in Education (Ofsted). 2005. *The National Literacy and Numeracy Strategies and the Primary Curriculum*. London: Ofsted nationalarchives.gov.uk/

Pedersen, S., A. Hill, and R. Callingham. 2015. “Construction and Validation of a Survey Instrument to Determine the Gender-related Challenges Faced by Pre-service Male Primary Teachers.” *International Journal of Research and Method in Education* 38 (2): 184-199. Taylor & Francis Online

Schmidt, W., M. Tatto, K. Bankov, S. Blömeke, T. Cedillo, L. Cogan, L., . . . L. Paine. 2007. *The Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries*: *Mathematics Teaching in the 21st Century*. East Lansing, MI: Center for Research in Mathematics and Science Education, Michigan State University. http://www.educ.msu.edu/content/sites/usteds/documents/MT21Report.pdf

Segall, A. 2004. “Revisiting Pedagogical Content Knowledge: The Pedagogy of Content/The Content of Pedagogy.” *Teaching and Teacher Education* 20 (5): 489-504. Science Direct

Senocak, E. 2009. “Development of an Instrument for Assessing Undergraduate Science Students’ Perceptions: The Problem-Based Learning Environment Inventory.” *Journal of Science Education and Technology* 18 (6): 560-569. Springer

Shulman, L. 1986. “Those Who Understand: Knowledge Growth in Teaching.” *Educational Research*, 15 (2): 4-14. SAGE

Shulman, L. 1987. “Knowledge and Teaching: Foundations of the New Reform.” *Harvard Educational Review* 57 (1): 1-22. Google Scholar

Stemler, S. 2004. “A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability.” *Practical Assessment, Research and Evaluation* 9 (4): 1-19. Google Scholar

Tatto, M. T., J. Schwille, S. Senk, L. Ingvarson, R. Peck, and G. Rowley. 2008. *Teacher Education and Development Study in Mathematics (TEDS-M): Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics. Conceptual Framework*. East Lansing MI: Teacher Education and Development International Study Center, College of Education, Michigan State University. https://msu.edu/~mttatto/documents/TEDS_FrameworkFinal.pdf

TEMAG [Teacher Education Ministerial Advisory Group]. 2014. *Action Now: Classroom Ready Teachers*. Canberra, Australia: Australian Government Department of Education and Training. Google Scholar

Withrow, B. 2013. *Research Methods in Crime and Justice*. New York, NY: Routledge. Taylor & Francis Group

Wright, B., and M. Stone. 1979. *Best Test Design. Rasch Measurement*. Chicago, IL: MESA Press. Google Scholar

Wright, B., and M. Stone. 1999. Fit Analysis. In *Measurement Essentials: 2nd Edition*. Wilmington, Delaware: Wide Range Inc. https://www.rasch.org

Zubairi, A., and N. Kassim. 2006. “Classical and Rasch Analyses of Dichotomously Scored Reading Comprehension Test Items.” *Malaysian Journal of ELT Research* 2 (1): 1-20. Google Scholar

Appendix A: Additional Mathematics Pedagogical Content Knowledge Instrument Items

<table>
<tr><td colspan="4">In the number sentence, 17 - 9 = ☐ + 1, one of your students places an 8 in the empty box. For each teacher intervention provided in the table below, indicate to the right which intervention you <i>would not use</i>, <i>might use</i>, or <i>definitely would use</i> to help the student understand this relationship.</td></tr>
<tr><td></td><td>Would NOT use</td><td>Might use</td><td>Definitely would use</td></tr>
<tr><td>Discuss with the student the purpose of the equal sign and about relationships between the left side and the right side of an equation.</td><td>[ ]</td><td>[ ]</td><td>[x]</td></tr>
<tr><td>Remind the student that what you do to one side of the equation you must do to the other side.</td><td>[x]</td><td>[ ]</td><td>[ ]</td></tr>
<tr><td>Question 30.3
Ask the student to solve a similar, yet less difficult problem such as 7 - 5 = ☐ + 1</td><td>[ ]</td><td>[x]</td><td>[ ]</td></tr>
<tr><td>Provide the student with a balance scale and blocks to create a representation of the equation.</td><td>[ ]</td><td>[ ]</td><td>[x]</td></tr>
<tr><td>Advise the student to consider the commutative property of addition.</td><td>[x]</td><td>[ ]</td><td>[ ]</td></tr>
</table>