Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry

Bette Davidowitz[a] and Marietjie Potgieter[b]

[a]Department of Chemistry, University of Cape Town, Cape Town, South Africa
[b]Department of Chemistry, University of Pretoria, Pretoria, South Africa

ABSTRACT

Research has shown that a high level of content knowledge (CK) is necessary but not sufficient to develop the special knowledge base of expert teachers known as pedagogical content knowledge (PCK). This study contributes towards research to quantify the relationship between CK and PCK in science. In order to determine the proportion of the variance in PCK accounted for by the variance in CK, instruments are required which are valid and reliable as well as being unidimensional to measure person abilities for CK and PCK. An instrument consisting of two paper-and-pencil tests was designed to assess Grade 12 teachers' CK and PCK in organic chemistry. We used the Rasch measurement model to convert raw score data into interval measures and to provide empirical evidence for the validity, reliability and unidimensionality of the tests. The correlation between CK and PCK was estimated as $r = .66$ ($p < .001$). We found evidence to suggest that while topic-specific PCK (TSPCK) develops with increasing teaching experience, high levels of CK can be acquired with limited teaching experience. These findings support the hypothesis that CK is a requirement for the development of TSPCK; proficiency in CK is, however, not necessarily associated with high levels of TSPCK.

KEYWORDS

Content knowledge; topic-specific pedagogical content knowledge; Rasch measurement model

Introduction

Research has shown that a high level of Content Knowledge (CK) is necessary but not sufficient for the development of the special knowledge base of expert teachers known as Pedagogical Content Knowledge (PCK) (Pitjeng, 2014, Rollnick, Bennett, Rhemtula, Dharsey & Ndlovu, 2008). In a review of pedagogical content knowledge in science education, Kind (2009) noted that possession of good CK is regarded as a pre-requisite for developing science teachers' PCK. Several researchers have carried out studies to quantify the relationship between CK and PCK in science subjects, for example Gramzow, Riese and Reinhold (2013) in physics, Tepner and Dollny (2014), Mavhunga and Rollnick (2013) and Rollnick and Mavhunga (2014) in chemistry, and Jüttner, Boone, Park and Neuhaus (2013) in biology. In statistical terms, this puzzle translates to the question: what proportion of the variance in PCK can be accounted for by the variance in CK? In order to begin to address this question one needs instruments which are valid and reliable, as well as being unidimensional in order to measure both variables. In addition, for a meaningful estimation of the correlation between CK and PCK the data collected using these instruments should be interval rather than ordinal data where intervals between data points have the same meaning irrespective of their position on the scale. The Rasch measurement model (Rasch, 1980) can address all of these

CONTACT Bette Davidowitz, bette.davidowitz@uct.ac.za

requirements in a convincing manner. Rasch analysis of raw score data can provide empirical evidence for validity, reliability and unidimensionality of instruments. During Rasch analysis ordinal raw score data are transformed into interval measures, which improves the rigour of the analysis and generates linear scores that are more suitable for correlation studies.

Boone, Townshend and Staver (2011) have emphasized the need for both sound theory and rigorous instrumentation to advance science education research. The theory of basic organic chemistry, how it is formulated and communicated, has been refined over more than a century, but the theory of PCK is still being developed. During October 2012 a group of researchers gathered in Colorado Springs for a summit on PCK to explore current and future research dimensions and to reach a consensus on PCK. They formulated the model of Teacher professional knowledge and skills (Gess-Newsome, 2015) to uncover the link between teachers' professional knowledge, topic-specific professional knowledge, classroom practice and student outcomes.

The basic tenets of this theory have been operationalised into instruments probing Topic-Specific PCK (TSPCK) in several disciplines, including chemistry. This study forms part of a larger project designed to develop and refine instruments probing Topic-Specific PCK in chemistry and physics. We have developed an instrument consisting of two paper-and-pencil tests for the assessment of content knowledge in foundational organic chemistry and teachers' PCK for teaching this topic to senior secondary school students. For this study we used the term content knowledge to mean 'the academic content of the discipline' which includes 'the science and engineering practices used to generate knowledge, the disciplinary core ideas, and the recognition of cross-cutting concepts.' (Gess-Newsome, 2015, p32). The design of the instrument will be presented briefly as a full account has been published elsewhere (Davidowitz & Vokwana, 2014). In this paper we report on the application of the Rasch model to validate and improve the rigour of the instrument for future use and for the special purpose of quantifying the relationship between CK and TSPCK.

The choice of the topic of organic chemistry was based on the introduction of the National Curriculum Statement in South Africa in 2006 (Department of Education, 2006) which brought about a shift in focus in the topic of organic chemistry for Grades 10-12. Students would be expected to understand concepts in organic chemistry as well as the relationship between molecular structure and physical properties of classes of organic compounds. The implementation of the NCS curriculum led to a greater emphasis on the assessment of this topic in the final examination at the end of Grade 12. Questions based on organic chemistry concepts, which formerly comprised around 11% of the Grade 12 final examinations currently account for over 30% of the chemistry paper. In addition to the curriculum and assessment changes described above, organic chemistry has long been regarded as difficult to teach and learn at both secondary and tertiary levels (Green & Rollnick, 2006; Hart, 1925; Katz, 1996). These factors created the need for the development of high quality tools to measure both content knowledge and PCK of teachers in this content area.

Literature review

Pedagogical content knowledge

In 1986, Shulman described Pedagogical Content Knowledge as the transformation of Content Knowledge into various forms which can be used to help students to understand the concepts. In a later paper he described pedagogical content knowledge as being:

> … of special interest because it identifies the distinctive bodies of knowledge for teaching. It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to diverse interests and abilities of learners, and presented for instruction. Pedagogical content knowledge is the category most likely to distinguish the understanding of the content specialist from that of the pedagogue." (Shulman, 1987, p. 8).

Many researchers have proposed models to characterise the transformation of pure CK to content knowledge for teaching as manifested in the classroom. For example, Cochran, DeRuiter and King (1993) proposed a model of pedagogical content knowing, PCKg, comprising "a teacher's integrated understanding of four components of pedagogy, subject matter content, student characteristics, and the environmental context of learning" (p. 266). Their model depicts a dynamic process in which PCKg continues to develop throughout a teacher's career and which enables teachers to "create teaching strategies for teaching specific content in a discipline" (p. 266) for example in organic chemistry. Geddis and Wood (1997) extended Shulman's (1986) idea of pedagogical content knowledge and focused particularly on the relationship between CK and PCK, elaborating on the transformation of CK which emerged from four components namely: students' prior concepts, subject matter representations, instructional strategies, and curriculum materials and curricular saliency. These authors use the term curricular saliency to refer to teachers' understanding of the place of a topic in the curriculum and the purpose(s) in teaching, their decisions to leave out certain aspects of a topic and their awareness of how a topic fits into the curriculum.

Based on the work of these researchers, Rollnick et al. (2008) developed a model which shows the integration of teachers' internal knowledge domains to produce the visible product of integration of these domains in the classroom, which they refer to as manifestations. This model is useful in that it separates the teacher's internal thought processes from what can be observed directly in the classroom as it allows the distillation of the overall teaching strategy produced in action, which is informed by the teacher's knowledge domains. The Rollnick et al. (2008) model was later refined to include the centrality of teachers' beliefs and their influence on the teachers' knowledge domains and vice versa, and, consequently, what is enacted in the classroom (Davidowitz & Rollnick, 2011).

Several authors (Loughran, Mulhall & Berry, 2004; Magnusson, Krajcik & Borko, 1999; Park & Chen, 2012; Van Driel, Verloop & de Vos, 1998 and Veal & MaKinster, 1999) have noted the topic-specific nature of PCK. More recently, Mavhunga and Rollnick (2013) defined and described Topic-Specific PCK (TSPCK) as the capacity to transform the subject matter of a given topic for the purpose of teaching. TSPCK is also related to Ball, Thames and Phelps' (2008) concept of specialised content knowledge for teaching in mathematics. Mavhunga and Rollnick (2013) extended the Davidowitz and Rollnick (2011) model to include TSPCK which results from the transformation of CK; thus content knowledge is considered to be a prerequisite to develop PCK.

TSPCK comprises teachers' understanding of:

- Students' Prior Knowledge
- Curricular Saliency (deciding what is important for teaching & sequencing)
- What makes a topic easy or difficult to teach
- Representations, including powerful examples & analogies
- Conceptual Teaching Strategies

Figure 1 shows the model of TSPCK which informed this study. The left hand side of the model is based on the internal domains of teacher knowledge in the Rollnick et al. (2008) model, while the right hand side describes the construct of TSPCK and illustrates how TSPCK results from the transformation of CK.

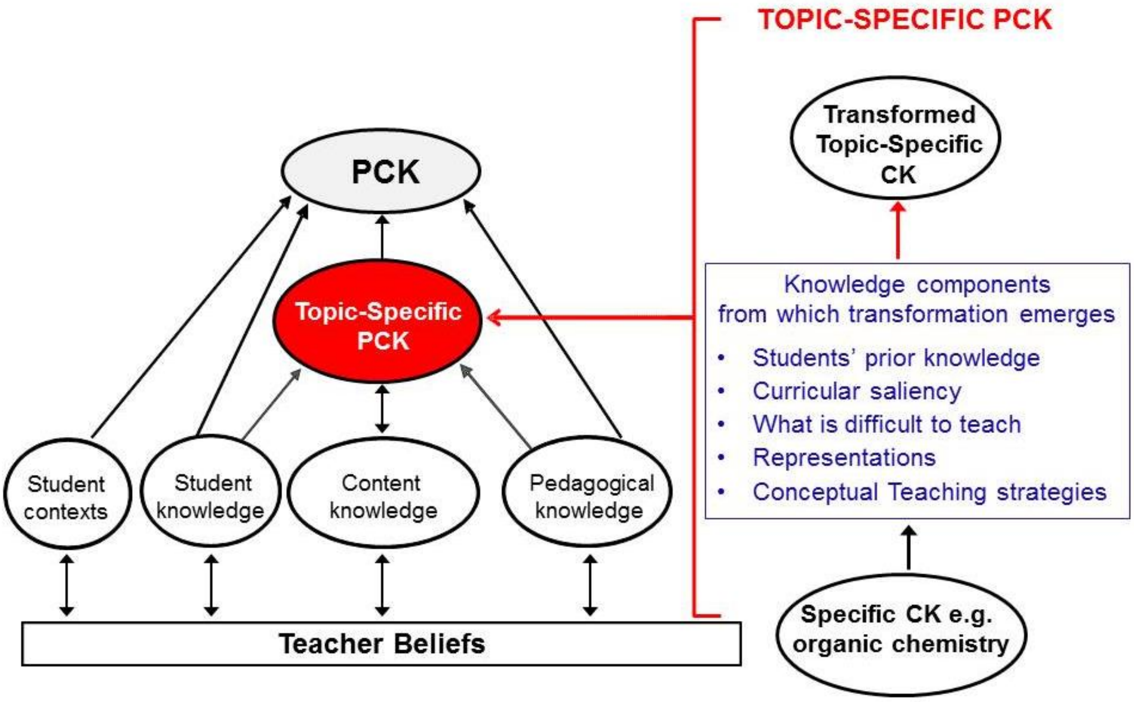


Figure 1. A model of TSPCK based on Mavhunga and Rollnick (2013)

Loughran et al. (2004) developed useful tools for capturing and portraying PCK which they called Pedagogical and Professional experience Repertoires (PaP-eRs) and Content Representations (CoRes). The latter lead to the identification of Big Ideas for teaching particular topics, while PaP-eRs are narrative accounts of practice designed to bring to life the ideas in the CoRe. The Big Ideas highlight a number of key concepts commonly viewed by practitioners as important for students to learn in order to understand the topic (Loughran et al., 2004). Whilst models of PCK and tools such as CoRes and PaP-eRS allow researchers to identify PCK in their research data, the process is time-consuming as most of the research is based on case studies. In addition, it is not possible to generalize the data to large cohorts of teachers in a specific context.

The model in figure 1 is based on the assumption that CK is a necessary prerequisite for TSPCK. Our study seeks to test this assumption. In order to do this we needed the appropriate tools. While assessment practise in education routinely seeks to capture proficiency in CK, the same is not true for TSPCK. One option would be to use quantitative data collection methods such as tests and questionnaires to assess and measure teachers' CK and PCK. A number of studies using specially designed instruments which attempt to measure teachers' PCK have been reported in the literature. These include work by Rohaan, Taconis, and Jochems (2011) in technology education, Gramzow et al. (2013) in physics and Jüttner et al. (2013) in biology. Similar studies in chemistry have focussed on foundational topics (Tepner & Dollny, 2014), chemical equilibrium (Mavhunga & Rollnick, 2013) and electrochemistry (Rollnick & Mavhunga, 2014). None of the chemistry tools were, however, specific with respect to organic chemistry, thus the first objective of the study was to design tests to probe both CK and TSPCK which would be appropriate for this particular content topic.

The Rasch measurement model

The Rasch measurement model (Andrich, 1988; Bond & Fox, 2007; Wright & Stone, 1999) elevates test development and utilisation to a level of sophistication that cannot be achieved when using raw scores only. This method facilitates the computation of *linear* item measures to reflect the relative difficulty of test items and *linear* person measures relating to the ability a person exhibits on a particular construct, in this case proficiency in organic chemistry in the one test and teacher PCK for teaching this topic to senior secondary school students in the other. Applying the Rasch model can enhance the quality of a project at various levels. The rigour of analysis is increased and the assumption of linearity can be met – an assumption often ignored in the analysis of educational data (Boone & Rogan, 2005, Boone et al., 2011).

The Rasch model describes the interaction between persons and items with the elegant, yet simple mathematical equation, $ln(P_{ni}/(1- P_{ni})) = \beta_n - \delta_i$, where P_{ni} is the probability of person n with ability β_n correctly answering item i with difficulty δ_i. The two variables, β_n and δ_i, are linear interval measures with the same log-odds units, called logits. This means that so called person-item maps, or Wright maps, can be constructed in which both person measures and item measures are plotted along the same axis. Wright maps facilitate a qualitative evaluation of a test instrument in terms of the match between the range of item difficulties and the range of performance abilities of the respondents, as well as potential item redundancy or inadequate cover within specific regions of task difficulty on the scale. Measurement error can be minimised when there is good alignment with person proficiencies and a good spread of the items across difficulty levels.

The development of a quality instrument that probes a specific trait in a comprehensive manner requires both qualitative and quantitative thought and reflection (Boone et al., 2011). The researcher must have a thorough understanding of the theory underpinning the trait and its expected manifestations so that varying strengths of the trait can be captured on a scale of "less than" to "more than" along a single dimension. Data that fits the model constitutes empirical evidence that the basic assumptions of measurement have been met and that meaningful inferences can be made from estimates of person ability and item difficulty. Various fit statistics are routinely evaluated during Rasch analysis, such as item fit, person fit and the overall item-trait interaction expressed as chi-square statistics. Polytomous items are also checked for threshold ordering which reveals whether item categories are functioning as expected (Linacre, 2002).

Another assumption of the Rasch model is that of local independence, which requires that every item contributes related, but independent information regarding the variable being measured. When violations of local independence occur, the reliability estimate of the test instrument is inflated giving a more favourable indication of internal consistency than is justified (Marais & Andrich, 2008a). We have used the Person Separation Index (PSI) to evaluate the internal consistency of our instrument, a statistic which is similar to the person reliability and item reliability indices reported in Winsteps. The derivation of the PSI is identical to that of the Cronbach's alpha from classical test theory, but it is computed from estimated person measures whereas Cronbach's alpha is calculated from non-linear raw scores (Clauser & Linacre, 1999). In the absence of evidence for local dependence between test items, the value of the PSI can be taken as a trustworthy estimate of the internal consistency of the instrument. The values for PSI range from 0 to 1 and this statistic is interpreted in a similar manner as Cronbach's alpha.

Aim of the study

This study attempts to evaluate both the CK and TSPCK of Grade 12 teachers in organic chemistry and to quantify the relationship between these two variables. We also investigated whether there is evidence for the development of CK and TSPCK with teaching experience.

Methodology

The starting point for this study was the design of suitable CK and TSPCK paper-and-pencil tests as independent entities which would be administered as a single instrument for data collection purposes. The design criteria, overall structure and the scoring guides for these tests have been described elsewhere (Vokwana, 2013, Davidowitz & Vokwana, 2014), thus only extracts from the two tests will be presented in this paper. The tests were piloted and revised based on the pilot results as well as feedback from expert teachers. Both tests comprised of five tasks, each consisting of a number of separate open-ended items, but with no explicit alignment between the tasks or the items.

The design of the CK test was informed by the syllabus for Grade 12 as well as by textbooks endorsed by the Department of Basic Education in South Africa. The National Curriculum Statement (NCS) for organic chemistry includes the following topics in organic chemistry namely functional groups, saturated and unsaturated structures, isomers, naming and formulae, physical properties, and chemical reactions (substitution, addition and elimination) (Department of Education, 2006). The test included tasks on molecular structures, generating and naming isomers, functional groups, types of reactions, and the relationship between molecular structures, intermolecular forces and physical properties. While nomenclature comprises an important part of the syllabus it was not tested separately, but was integrated into the other tasks. The items in the CK test were classified according to the knowledge dimensions proposed by Jüttner et al. (2013), namely declarative knowledge – knowing facts and other information appearing in texts, procedural knowledge - knowing how to execute a skill or apply concepts and principles to specific questions in organic chemistry, and conditional knowledge - knowing when and why to use declarative or procedural knowledge, for example to generate formulae of isomers or to explain trends in boiling points of alcohols. Table 1 shows the classification of the items in the CK test according to topics and knowledge types. There was some ambiguity in the categorisation of components of tasks 3 and 4 as either declarative or procedural knowledge, because some teachers may have been so familiar with the tasks that they could retrieve the answers from memory.

Table 1. Classification of tasks in the CK test according to Jüttner et al. (2013)

	Topics				
Knowledge dimensions	Drawing structures	Drawing and naming isomers	Determining product of reaction	Identifying functional groups	Relationship between intermolecular forces and boiling points of organic compounds
Declarative				4 items, Q4a-d	
Procedural	2 items, Q1	2 items, Q2a	6 items, Q3	2 items, Q4e and Q4f	2 items, Q5a and Q5c
Conditional		4 items, Q2b and Q2c			1 item, Q5b

The content of the test was validated by a group of four expert teachers from local high schools. Teachers' responses on the CK test were scored as for a conventional test giving partial credit where appropriate. A typical example of a task is included in Figure 2.

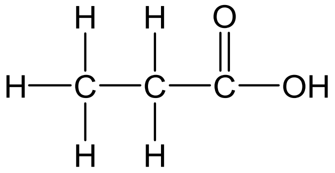
2. Molecules with the same molecular formula can have different structures, this is known as isomerism. There are several kinds of isomerism such as structural and geometric. Answer the following question about isomers.

a) Draw the following isomers for 2-butene.

Structural isomer of 2-butene

Geometric isomer of 2-butene

b) Structural isomers can belong to different homologous series; in other words they could have different functional groups. Draw two structural isomers with different functional groups for the compound below.



$$\mathrm{H{-}\underset{H}{\overset{H}{C}}{-}\underset{H}{\overset{H}{C}}{-}\overset{O}{\overset{\|}{C}}{-}OH}$$

Structural isomer of propanoic acid

(Item 2bi)

Structural isomer of propanoic acid

(Item 2bii)

c) IUPAC name of isomer above

(Item 2ci)

IUPAC name of isomer above

(Item 2cii)

Figure 2. CK item probing the ability to generate isomers from a given structure

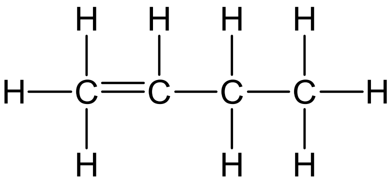
The design of the TSPCK test posed a particular challenge as we had to operationalise the five components of the Mavhunga and Rollnick (2013) model of TSPCK for foundation level organic chemistry, something which has not been achieved to date. In addition, the emphasis during the development of the TSPCK test was to ensure that it was fit for purpose, i.e. to formulate tasks that would make explicit teachers' tacit TSPCK. Teachers are unlikely to have seen any items like the tasks formulated in this test, thus it was not possible to predict what their responses might be and the level of challenge posed by the tasks. The TSPCK test consisted of five sections corresponding to the five categories of TSPCK (Mavhunga & Rollnick, 2013), namely students' prior knowledge, curricular saliency, what makes a topic easy or difficult to teach, representations and conceptual teaching strategies. As for the CK test, items were open-ended and included tasks based on concepts in the Grade 12 syllabus, for example drawing and naming molecular structures, recognition of functional groups and the use of representations of molecules to teach the relationship between intermolecular forces and boiling points of alkanes.

Since there are a wide variety of responses possible for the TSPCK test these were scored using a rubric. For example, Park, Jang, Chen and Jung (2011) used a rubric to investigate the correlation between a teacher's level of PCK and the degree to which his/her classroom is reform-oriented. Mavhunga and Rollnick (2013) constructed a rubric based on their model of TSPCK and used it to score teachers' responses for the topic of chemical equilibrium. This rubric was adapted to score teachers' responses to the TSPCK test for organic chemistry. Teachers' responses were scored on a four-point scale ranging from 1 (Limited), 2 (Basic), 3 (Developing) to 4 (Exemplary) TSPCK for each task (Davidowitz & Vokwana, 2014). A zero was assigned if the teacher failed to provide a response to the task. Figure 3 shows an example of a task probing teachers' conceptual teaching strategies and presents part of the rubric used to score it.

Scores were peer validated by independent raters and an interrater reliability analysis using the Kappa statistic was carried out using Medcalc version 15.11.0 (https://www.medcalc.org; 2015) to determine consistency among raters. The value of Fleiss' Kappa is 0.82 ($p<0.001$), 95% CI 0.779 to 0.856 which is considered to be very good agreement between raters. An example of a response to item E2 which was coded as exemplary is given below.

> "I'd ask them to draw C's structure in various ways (e.g. Aufbau & Lewis dot), then ask how many valence e's (electrons) a C atom has [4], then how many bonds each C can form [4] & therefore what was wrong [one of the Cs has 5 bonds]. Any incorrect answer in this dialogue would alert me to other misunderstanding I'd need to help the child with." Teacher 20, TSPCK test.

E2 In a diagnostic test a student drew the incorrect structure below.



$$\begin{array}{ccccccccccc} & & H & & H & & H & & H & & \\ & & | & & | & & | & & | & & \\ H & — & C & = & C & — & C & — & C & — & H \\ & & | & & & & | & & | & & \\ & & H & & & & H & & H & & \end{array}$$

Given that you have taught your students how to draw structural formulae, how would you conduct a revision lesson, to correct this student's response?

TSPCK Components	**(1) Limited**	**(2) Basic**	**(3) Developing**	**(4) Exemplary**
Teaching Strategies Section E	Chooses strategy that • Gives only the correct answer to the task • No confirmation or confrontation of student prior knowledge and/or common misconceptions	Chooses strategy that • Considers confirmation/ confrontation of student prior knowledge and/or common misconceptions • Uses only symbolic representation • No conceptual approach	Chooses strategy that • Considers confirmation and confrontation of student prior knowledge and/or common misconceptions • Conceptual approach to topic not clear • Uses a representation or model of a molecule to enforce an aspect of a concept.	Chooses strategy that • Considers confirmation and confrontation of student prior knowledge and/or common misconceptions • Uses a model or symbolic representation to enforce a singular aspect of a concept. • Conceptual approach explicit

Figure 3. A task from the TSPCK questionnaire, E2, and the section of the rubric used to score teachers' responses for item E2

Sample

The tests were completed by a cohort of teachers from diverse schools across South Africa, N = 89. This was an opportunistic sample as we had to persuade teachers to complete the tests in addition to the demands of their teaching. While the sample represents the full range of teacher competencies, it does not reflect the distribution of abilities to teach organic chemistry across the teaching profession in South Africa. Just over half the sample of teachers was female (52.3%). Their teaching experience ranged from 1 to 45 years with 52.8% of teachers having taught for 10 or more years. There was a great diversity in teacher training with teachers having studied at teacher training colleges or universities to complete a variety of degrees and/or diplomas. Of those teachers who studied at universities, some would have completed a science degree which included courses in chemistry followed by a post graduate diploma in teaching while others would have been enrolled in a bachelor of education degree. A breakdown of the sample in terms of years of experience is provided in Table 2. The categories shown are based on Schneider and Plasman's (2011) categorisation of teacher experience.

Table 2. Sample of teachers participating in the study

Category	**Years of teaching**	**Number of teachers, %***
Novice	< 5 years	17 (19.1)
Early career	5-9 years	24 (27.0)
Mid-career	10 - 18	22 (24.7)
Late career	> 18 years	25 (28.1)

* Percentages do not add to 100 since there was information missing from one data record.

Data analysis and findings

In this section the Rasch analysis and subsequent refinement of the CK and TSPCK data sets will be described. *Post hoc* refinement of the data was informed by a joint consideration of the empirical findings and a qualitative appraisal of the content of test items and the way they were scored. The Rasch analysis of the data was done using the RUMM2030 suite of software programmes (Andrich, Sheridan & Luo, 2011). According to Tennant and Conaghan (2007) the performance requirements for a test instrument of high quality are the following: empirical evidence for unidimensionality; acceptable fit statistics for both items and persons, i.e. the extent to which persons answered the items in a consistent and predictable manner; proper targeting of the test with a good spread of item difficulties across the range of person abilities; and an acceptable estimate of the internal consistency or reliability of the instrument. In order to ensure that all items were working as expected they were carefully checked for fit, discrimination, appropriate functioning of response categories, and the absence of both response independence and differential item functioning (DIF) for subsets of respondents. There are rich resources available that provide a detailed exposition of these test and item characteristics and their interpretation, for example Bond and Fox (2007), Boone, Staver and Yale (2014), Smith and Smith (2004) and the comprehensive website maintained by the Institute for Objective Measurement, Inc. (http://www.rasch.org/).

Missing data

The CK test consisted of 23 items and the TSPCK instrument of 10 items. There was no missing data in the CK data set, but the TSPCK data set had a 9.3% prevalence of missing responses (83 of 890) which occurred more frequently for items that were anticipated to be more demanding (B3, D2 and E2, 15% missing responses). We therefore decided to score missing responses as 0 (zero), based on the assumption that teachers were unable to answer these questions because they lacked the appropriate knowledge or could not make their knowledge explicit.

Analysis of CK data

The CK instrument consisted of 19 dichotomous and four polytomous items, with the polytomous items scored on a scale of increasing competence. The partial credit parameterisation of the Rasch model (Masters, 1982) was therefore chosen for the analysis of the data. The data showed acceptable item and person fit statistics. Only one item had a fit residual marginally outside the boundary of ±2.5, but after Bonferroni adjustment the probability of its occurrence was not significant. (Fit residuals in RUMM are equivalent to OUTFIT ZSTD in Winsteps.) This means that all of the items are working consistently together to measure the same construct. The polytomous item, Q5a, had to be rescored because one of its categories was not working as expected (Linacre, 2002). The thresholds between categories for this item were disordered and the plot of category probabilities indicated that there was no region on the continuum of person abilities where a score of 3 out of a maximum of 4 was most likely. After careful consideration of the scoring guide this situation was addressed by combining the categories for scores of 3 and 4, thereby reducing the maximum score for the item to 3 (Boone, et al., 2014, pp. 191-216). Two sets of items showed response dependence, items Q2bii and Q2cii which were related to the same compound (Figure 2), and items Q4b and Q4c, which required the recognition of a ketone and an aldehyde, respectively, structures which are closely related. This situation was resolved by the formation of higher-order polytomous items (subtests), Q2bcii and Q4bc, in which the inflationary effect of the dependence is absorbed (Marais & Andrich, 2008b).

The data did not show any differential item functioning based on the gender or level of experience of the respondents. A t-test was performed guided by a Principal Component Analysis of Residuals as proposed by Smith (2002) which provided empirical evidence for the unidimensionality of the data. This finding also provides empirical evidence for the construct validity of the instrument when interpreted in conjunction with the qualitative judgement of expert teachers. The data fitted the model well (Total item chi-square 54.26, df 42, prob 0.0972) and the Person Separation Index for the instrument was good, PSI 0.81. The person-item (or Wright) map for the CK test is shown in Figure 4 below.

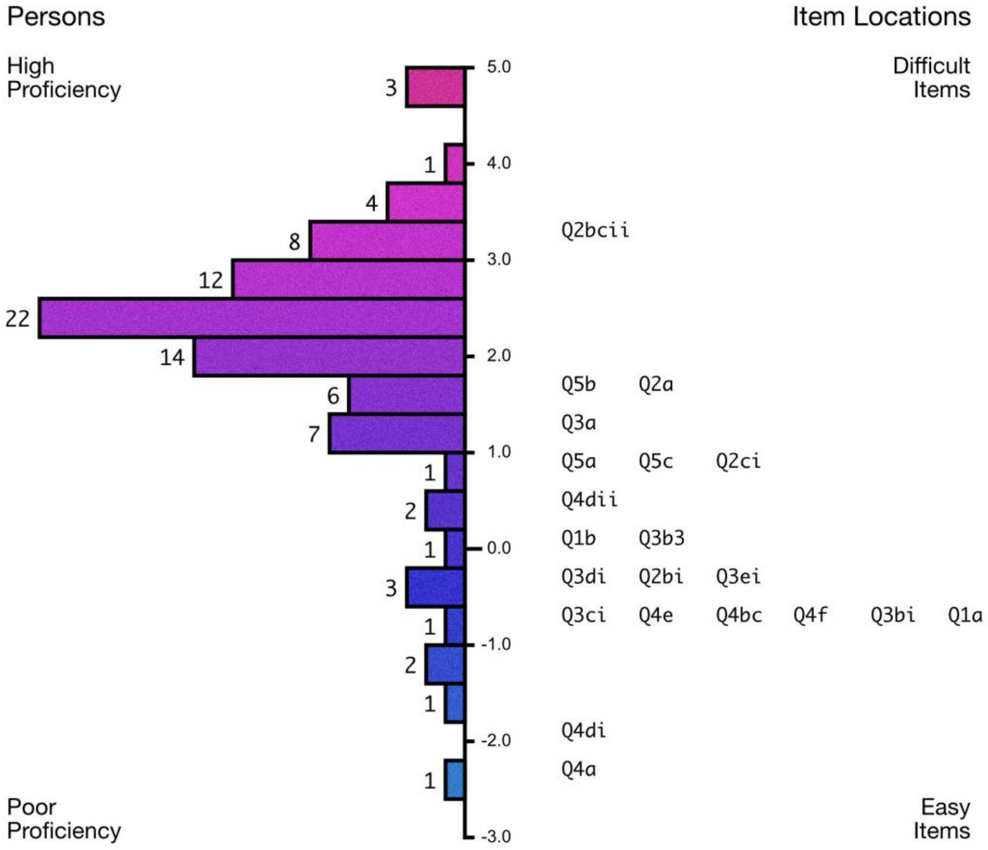


Figure 4 Wright map for the CK questionnaire, N = 89. Mean person performance = 2.08 (std. dev. 1.31)
Note: Teachers and items are located on the vertical scale according to their respective person ability and level of item difficulty, respectively. The number of teachers with a specific proficiency is indicated to the left of each bar.

The item mean in the person-item map is automatically set at zero during the analysis and person proficiencies are estimated relative to this mean. Person proficiencies and item difficulties are located on the same scale with poor proficiency and easy items at the bottom of the map and high proficiency and difficult items at the top. The distribution on the left of the vertical axis shows the spread of performance scores of the teachers relative to the spread of item difficulties to the right of the same axis. The specific location of a person on the map indicates that for this person the probability of providing a correct response to dichotomous items of matching difficulty on the same vertical scale is 50%. It is higher than 50% for dichotomous items of lower difficulty and lower for items of higher difficulty. In this instrument all items except Q2a, Q5a, Q5b and Q5c are dichotomous where the answer is either right or wrong. Thus in Figure 4, the bar on the left labelled "2" at ca. δ 0.5 logits indicates that these 2 teachers had a ca. 50% probability of giving the correct answer for Q4dii and a more than 50% probability of giving the correct answers to dichotomous items below this level on the scale. The interpretation of the Wright map for polytomous items is more complex. The location of a polytomous item is at its mean calibration, i.e. the location at which a score in the top and bottom categories for that item are equally probable.

The person-item map in Figure 4 is a rich source of information that should be explored further. The mean person measure of 2.08 (standard deviation, 1.31) reflects the fact that most teachers in the sample experienced the test as being easy. Three teachers achieved a perfect score for this test (this is noted at the top left side of the Wright map with the notation 3 near the logit value of 5.0). However, the number of teachers with poor performance extending to *ca.* -2.5 logits at the bottom left side of the map is a cause for concern. The distribution of item difficulties in Figure 4 align with the categorisation of items in Table 1, with items requiring declarative knowledge easier than procedural items and items requiring conditional knowledge the most challenging. A similar outcome was obtained by Jüttner et al. (2013) which suggests that mastery of content develops from a low level of sophistication (declarative knowledge) to intermediate sophistication (procedural knowledge) and finally to the highest level of sophistication, namely conditional knowledge.

The majority of teachers in this cohort found generating structural and geometric isomers challenging (task 2, Figure 2). Task 2b required teachers to draw structural isomers of propanoic acid; each isomer had to contain a different functional group. Table 3 below shows the pattern of responses for this item. The *generation* of the first isomer (Q2bi, δ -0.42) was more difficult than the task of *selecting* a structural isomer from a group of potential answers (Q4f, δ -0.75). Once the structure required in item Q2bi was produced, the naming of this compound posed a further challenge (Q2ci, δ 0.96). The generation and naming of the second isomer of propanoic acid was the most difficult task in the test (Q2bcii, δ = 3.31). Despite explicit instructions in the problem statement, more than half the cohort of teachers (52.8%) drew two esters as their responses to this question. Answering questions on boiling points (Q5a-c) proved to be challenging for some teachers where they would have to draw on and apply the concepts of intermolecular forces in order to provide the correct answer.

Table 3. Pattern of responses to items 2b and c

Question	Teachers' responses	Percentage of correct answers
2bi	Ester: methyl ethanoate or ethyl methanoate	87.5
2bii	Hydroxyaldehyde, hydroxyketone, cyclic diol or enediol	18.0
2ci	Correct name for ester	73.9
2cii	Correct name for other isomer	11.2

It can reasonably be expected that expert teachers should have mastered at least the content knowledge that is prescribed in the syllabus for the students that they are teaching, but also more advanced material which could direct their teaching even though it may not be explicitly included in the syllabus. The only items that probed content knowledge beyond the prescribed syllabus were items Q2b and Q2c. On all the other items we would argue that a teacher should demonstrate a high proficiency, or in terms of the Rasch model, should have a high probability of answering correctly. This would be true of all teachers with a person measure of 1.40 and above (70 teachers, or 79% of our sample), but not for at least 12 teachers (13% of our sample) with an estimated person measure of 1.00 or lower. Teacher T57 (person measure -2.2) has clearly not mastered any of the content assessed by this instrument. The other eleven teachers with person measures below 1.00 should be able to recognise functional groups (task 4) and have developed a moderate command of the concepts probed in task 3 (reaction types). However, their command of task 5 (intermolecular forces and physical properties) would be marginal and they are unlikely to be able to teach students the insights needed to generate structural and geometric isomers of organic compounds (task 2). The person-item map shows a good spread of item difficulties across the scale up to item Q2a, but not beyond that. Item difficulties did not align well with person proficiencies which compromise the measurement precision that can be achieved with this instrument, especially for the more proficient teachers. In addition, there was a ceiling effect in the data because the instrument could not distinguish between the proficiencies of three teachers with perfect scores. The instrument could be further refined by reducing redundancy in the lower region of item measures around -1.0 logits and introducing items of higher demand which would allow more accurate estimation of the ability of higher performing teachers.

Analysis of TSPCK data

Scores were assigned to the ten items in the instrument with each being rated on a five point scale, from 0 (No response) to 4 (Exemplary) as shown in Figure 3. The partial credit parameterisation of the Rasch model (Masters, 1982) was used for the analysis of PCK data. The data fitted the model well, but four data records displayed misfit and were removed for the purpose of instrument evaluation. In addition, response categories for three of the items (B3, E1 and E2) were not working as expected. After careful evaluation of the empirical results together with the scoring guide, Item B3 was rescored by combining categories 1 and 2, E1 by combining categories 3 and 4, and Item E2 by collapsing categories 1 and 2, and 3 and 4. These *post hoc* refinements resolved all of the category disorder that was observed; however, for further refinement it will be important to consider whether the rubric was not specific enough to distinguish between categories for these items or whether it artificially forced four categories onto a task for which the data suggested fewer performance levels. None of the items displayed DIF based on gender or level of experience. The t-test analysis of item subsets provided no empirical evidence for multidimensionality. After *post hoc* refinement the fit of the data to the model was excellent (Total item chi-square 7.19, df 20, prob 0.996), and the estimates for the reliability of the instruments were very good (PSI 0.87). The map of person measures and item thresholds generated by Rasch analysis of the TSPCK test data is shown in Figure 5 below. The first threshold of item B1 (B1.1) is not shown on this map because it was off scale (-8.0 logits). The Wright map shown in Figure 5 should be interpreted as follows. Seven of the items (excluding B3, E1 and E2 for which categories were combined in *post hoc* refinement) had five scoring categories - zero (no response) or a range of 1 to 4 depending on the level of sophistication of the response. This means that there are four thresholds between adjacent categories for each of these items where the probability of a score of either 0 or 1, and 1 or 2, etc., are equally likely. These thresholds, or pivot points, between adjacent item categories of polytomous items are shown on the right hand side of the Wright map ranging from the lowest challenge (first threshold of the easiest item) (B1.1, -8.0 logits, not shown) to the highest challenge (highest thresholds of the most difficult items) at *ca.* 2.2 logits.

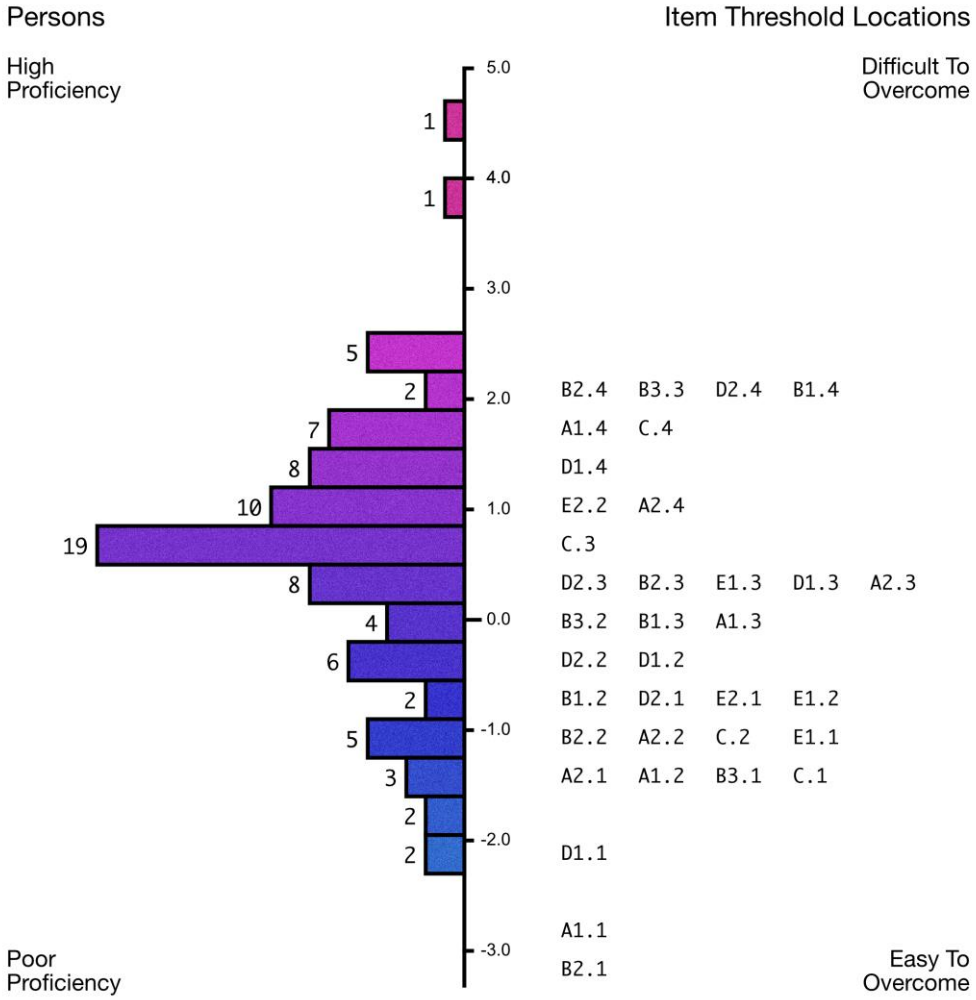


Figure 5. Wright threshold map for the TSPCK questionnaire, N = 85. Mean person performance = 0.63 (std. dev. 1.25)

Note: The threshold identity for specific items is indicated by the appropriate suffix appended to the item label. The threshold for B1.1 at -8.0 logits is not included.

The mean person performance was 0.63 (standard deviation 1.25). Estimated teacher abilities ranged from weak (-2.0 logits) to very strong with one teacher achieving a perfect score for this test shown at the top left side of the map between the scale values of 4.0 and 5.0. The person-item threshold map indicates that there is better alignment between item difficulties and person performance than in the CK test and the full complement of item response categories span the range of person proficiencies very well except for a few cases of top performing teachers. It is worth noting the finding that it was most difficult for teachers to demonstrate exemplary performance on the three tasks designed to probe curricular saliency, B1, B2 and B3, and D2, which probed the ability to use representations to explain concepts. Item B1 required the identification of foundational concepts for organic chemistry, B2 probed the identification of Big Ideas and their sequencing and item B3 assessed teachers' ability to formulate a coherent conceptual progression of the topic and insight into its importance and application. These findings suggest that the development of curricular saliency requires a more advanced knowledge of organic chemistry to enable teachers to formulate the Big Ideas and the order in which they should be sequenced.

Correlation between CK and TSPCK

According to the Mavhunga and Rollnick (2013) model of TSPCK, teachers with good CK are more likely to develop high levels of TSPCK while teachers with low levels of CK are likely to have low levels of TSPCK. Such a relationship is expected, as logic tends to suggest that with low CK there can be little TSPCK or as stated by Baumert, Kunter, Blum, Brunner, Voss, Jordan, Klusmann, Krauss, Neubrand and Tsai (2010) "PCK is inconceivable without a substantial level of CK" (p. 155).

A scatter plot of TSPCK person measures versus CK measures was constructed to explore the extent to which this is true for this cohort of teachers, see Figure 6 below.

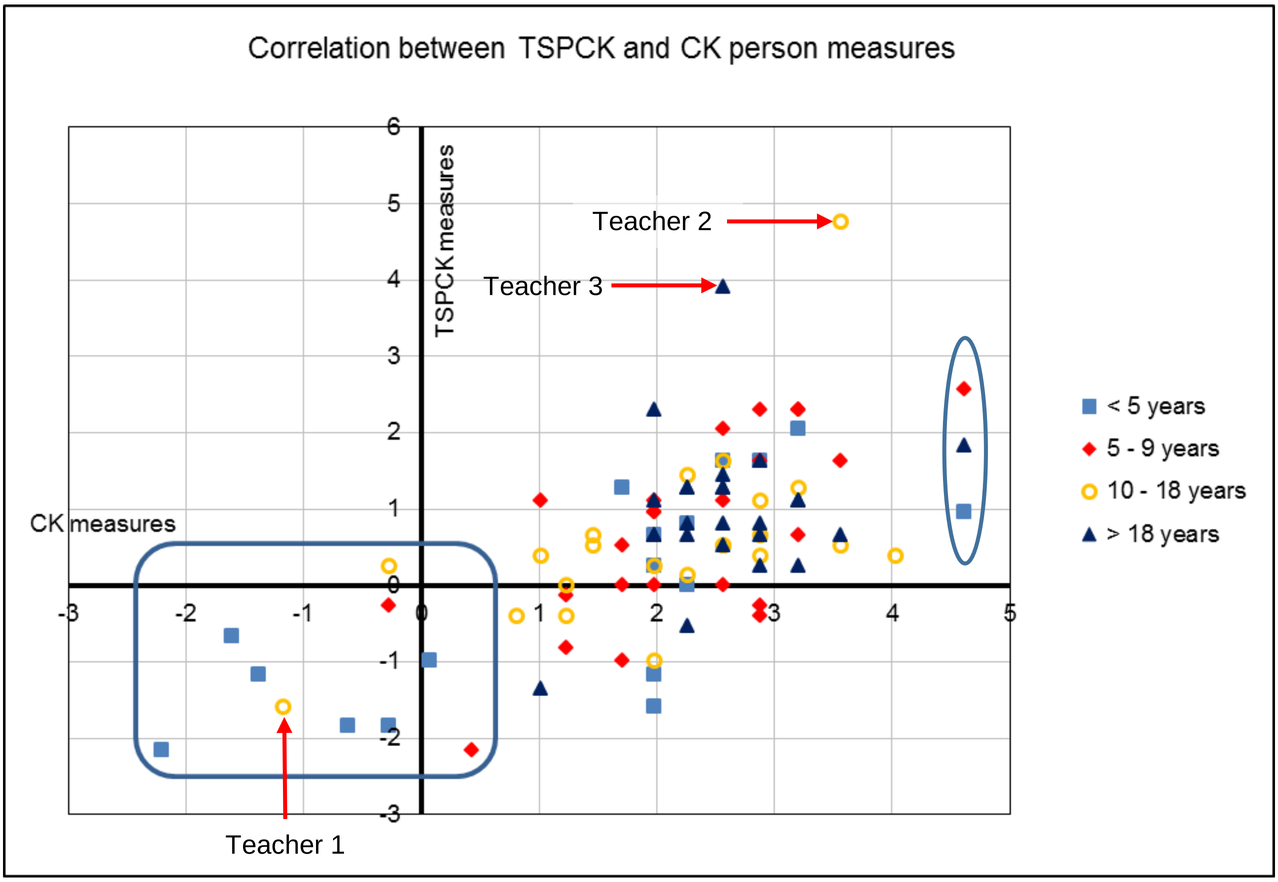


Figure 6. Scatter plot of person measures for CK and TSPCK, N=84, mean person performances: CK = 2.08 (std. dev. 1.31); TSPCK = 0.63 (std. dev. 1.25)

Note: Teacher Content Knowledge (CK) is represented on the horizontal axis from poor proficiency on the left to high proficiency on the right. Teacher Topic-Specific Pedagogical Content Knowledge (TSPCK) is presented on the vertical axis from poor proficiency at the bottom to high proficiency at the top.

CK measures are represented on the X-axis and TSPCK measures on the Y-axis. Different symbols were used to indicate the level of experience of the teachers. The values for five teachers were omitted from the correlation analysis, four because of the misfit of their PCK data and the fifth because of missing data about the level of experience of the teacher. We obtained a moderately strong sample correlation of 0.66 between teacher person measures for TSPCK and CK. This correlation coefficient was significantly greater than zero (one-sided p-value < 0.001, $N = 84$), indicating that TSPCK and CK are positively correlated in the population of teachers in this study.

Teaching experience as indicator for CK and TSPCK

We were also interested in looking for evidence of the development of teachers' CK and TSPCK with time as they gain experience in the classroom. The scatter plot in Figure 6 provides some information in this regard. There were ten teachers in our sample with very poor CK and TSPCK below the mean for the sample, shown in the box that is superimposed on the graph in Figure 6. Six of these teachers were novices with less than five years' experience, but surprisingly also present in this group were two early career and two mid-career teachers, such as teacher 1 who has been teaching for 12 years. As expected, the majority of novice teachers displayed below average CK and TSPCK whereas the majority of late career teachers (more than 18 years' experience) demonstrated CK and TSPCK above the mean for the sample. The picture for early and mid-career teachers was mixed with some teachers displaying excellent CK and TSPCK, while others such as those included in the box are clearly not adequately skilled for their task despite having been in the profession for many years. The scatter plot shows two outliers in terms of exemplary TSPCK (teachers 2 and 3); these teachers had 17.5 and 20 years' experience, respectively. Noteworthy is the fact that the three outliers in terms of excellent performance in CK shown in the oval in Figure 6 had either limited teaching experience (4 or 5 years) or extensive experience (30 years). This finding suggests that teachers may achieve excellence in CK very early in their careers or even before they enter the profession, whereas TSPCK seem to require extensive classroom experience to be perfected. On the other hand, it was clear for this sample that many years of classroom experience did not ensure that teachers perfected their craft, both in terms of CK and TSPCK.

Summary and conclusions

The Rasch analysis revealed that both tests met the performance criteria of good test design. Empirical evidence was obtained for the unidimensionality of both tests, items seem to define the trait well, there does not appear to be more than expected misfit and none of the test items showed DIF. *Post hoc* refinement was used to address issues of response dependence and the malfunctioning of response categories. The Person Separation Index for the CK test was 0.81 while that of the TSPCK test was found to be 0.87. There was room for improvement of both tests in terms of alignment with respondent proficiencies and the spread of item difficulties. In meeting the basic assumptions of the Rasch model as described above, we appear to have been successful in constructing a linear scale for the assessment of CK and TSPCK in foundational organic chemistry which can be used for data collection from large cohorts of teachers. A statistically significant correlation, $r = 0.66$ ($p < 0.001$, $N = 84$: see Figure 6) was found between the CK and TSPCK. This finding implies that an estimated 43.0% of the variance in TSPCK is accounted for by the variance in CK.

The correlation between CK and TSPCK shown above is stronger than the value of 0.36 reported by Tepner and Dollny (2014) using raw score data. Rollnick and Mavhunga (2014) reported a correlation of 0.54 for a study on electrochemistry in which the Rasch analysis was done for only one of the scales, namely the TSPCK data. Neither of these studies involved the transformation of both CK and TSPCK data to linear measures as recommended by Boone et al. (2011). Jüttner et al. (2013) used the Rasch method in order to estimate the correlation of biology teachers' CK and PCK to a higher level of precision. They reported a value of $r=0.22$ ($p \leq 0.01$, $N= 158$) for their study which is lower than the value obtained in this study. A possible reason for our finding of a higher correlation between CK and PCK could be that there was a greater similarity in the insights required to answer the CK and TSPCK questions on organic chemistry than was the case for these studies on other chemistry topics or biology.

Our findings show that low levels of CK are likely to be associated with lower levels of TSPCK while high levels of CK do not necessarily translate into high levels of PCK. As shown in Figure 6 teachers may achieve excellence in CK with limited professional experience, whereas TSPCK requires extensive classroom experience to be perfected. On the other hand, classroom experience alone is no guarantee for teachers to perfect their craft, both in terms of CK and TSPCK. The scatter plot of CK and TSPCK measures (Figure 6) provides some evidence for the gradual development of CK and TSPCK with teaching experience; however, since the sample sizes are small and the data on training and experience were self-reported, caution has to be exercised in interpreting the data. These findings invite further research to describe and analyse the pre-service training and professional environments of teachers at a similar stage of their careers but different levels of proficiency, such as teachers 1, 2 and 3, in an attempt to better understand the factors that did not ensure that teachers perfected their craft, in terms of both CK and TSPCK while in service.

Implications

As part of this study assessment tasks were developed to operationalise the theory of TSPCK for the topic of organic chemistry at the level at which this topic is taught in South African high schools. The Rasch model was applied to the data that was collected by means of these tasks to determine whether a meaningful scale was constructed for measurement. The results of the Rasch analysis indicated that this objective was achieved, something which has not been done before for the topic of organic chemistry. Both the CK and TSPCK tests met the requirements for good test design in terms of construct validity and reliability. The data fitted the Rasch measurement model well, which means that bold claims can be made about the transformation of total raw scores to linear person measures on an interval scale. However, there is still some room for improvement of both the CK and TSPCK tests with regards to the alignment of the test items with teacher proficiencies and in terms of the rubric that was used for the scoring of TSPCK data. It will also be important to collect data from a bigger and more representative sample of teachers in South Africa to confirm these findings and to monitor the consistency of the instruments over time.

Evidence was obtained for a linear relationship between CK and TSPCK and the strength of the relationship was quantified for the topic of organic chemistry. Correlation between CK and TSPCK was higher than that reported for similar studies in chemistry, for example Tepner and Dollny (2014) and Rollnick and Mavhunga (2014). Figure 6 reveals that the 10 teachers with the lowest CK also had weak TSPCK and were found in three of the four bands of teaching experience. There is also evidence to suggest that while TSPCK does seem to develop with increasing teaching experience; high levels of CK can be acquired with limited teaching experience. These findings support the hypothesis that CK is a requirement for the development of TSPCK; proficiency in CK is, however, not necessarily associated with high levels of TSPCK. Our study also highlights the need for the professional development throughout the span of a teaching career that focusses on both the CK and the TSPCK of the teacher. Development of either CK or TSPCK alone is likely to have limited success in terms of improving teacher effectiveness. Further work is needed to explore the contextual factors that enable or constrain development of TSPCK for in-service teachers so that appropriate support and professional development opportunities can be provide

Funding information

This work is based on the research supported by the National Research Foundation of South Africa [**g**rant **n**umber 90346]. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author(s) and therefore the NRF does not accept any liability in regard thereto.

References

Andrich, D. (1988). *Rasch Models for Measurement*. Beverly Hills, CA: Sage Publications.

Andrich, D., Sheridan, B., & Luo, G. (2011). *RUMM2030 software and manuals.* Perth: University of Western Australia.

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.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand M., & Tsai, Y-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal, 47*(1), 133-180.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Boone, W., & Rogan, J. (2005). Rigour in quantitative analysis: The promise of Rasch analysis techniques. *African Journal of Research in Science Mathematics and Technology Education, 9*(1), 25-38.

Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*, Dordrecht, Netherlands: Springer.

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch Theory to Guide the Practice of Survey Development and Survey Data Analysis in Science Education and to Inform Science Reform Efforts: An Exemplar Utilizing STEBI Self-Efficacy Data. *Science Education, 95*(2), 258-280.

Clauser, B. & Linacre, J. M. (1999). Relating Cronbach and Rasch Reliabilities. *Rasch Measurement Transactions, 13*(2), 696.

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical Content Knowing: An Integrative Model for Teacher Preparation. *Journal of Teacher Education, 44*(4), 263-272.

Davidowitz, B., & Rollnick, M. (2011). What lies at the heart of good undergraduate teaching? A case study in organic chemistry, *Chemistry Education Research and Practice*, *12*, 355–366.

Davidowitz, B., & Vokwana, N. (2014). Developing an instrument to assess Grade 12 teachers' TSPCK in organic chemistry. In H. Venkat, M. Rollnick, M. Askew and J. Loughran (Eds). *Exploring Mathematics and Science Teachers' Knowledge: Windows into teacher thinking*, (pp. 178-194). Oxford: Routledge.

Department of Education. (2006). *National curriculum statement grades 10-12 Physical Sciences content document*. Department of Basic Education. Pretoria, South Africa. Retrieved from http://wced.pgwc.gov.za/circulars/minutes07/NCS-PhysicalScience-Content-June06.pdf

Geddis, A. N., & Wood, E. (1997). Transforming subject matter and managing dilemmas: a case study in teacher education. *Teaching and Teacher Education, 13*(6), 611-626.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28-42). New York: Routledge.

Gramzow, Y., Riese, J., & Reinhold, P. (2013). *Prospective physics teachers' pedagogical content knowledge – validating a test instrument by using a think aloud study*. Paper presented at the European Science Education Research Association 2013 Conference, Nicosia, Cyprus. Retrieved from http://www.esera.org/media/esera2013/Yvonne_Gramzow_12Feb2014.pdf

Green, G., & Rollnick, M. (2006). The Role of Structure of the Discipline in Improving Student Understanding: The Case of Organic Chemistry. *Journal of Chemical Education, 83*(9), 1376-1381.

Hart, E. (1925). Teaching Organic Chemistry. *Journal of Chemical Education, 2*(2), 110-111.

Jüttner, M., Boone, W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK), *Educational Assessment Evaluation and Accountability, 25*, 45-67.

Katz, M. (1996). Teaching Organic Chemistry via Student-Directed Learning: A Technique that Promotes Independence and Responsibility in the Student. *Journal of Chemical Education, 73*(5), 440-445.

Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education, 45*(2), 169-204.

Linacre, M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Loughran, J., Mulhall, P., & Berry, A. (2004). In Search of Pedagogical Content Knowledge in Science: Developing Ways of Articulating and Documenting Professional Practice. *Journal of Research in Science Teaching, 41*(4), 370-391.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-132). Dordrecht, The Netherlands: Kluwer Academic.

Marais, I., & Andrich, D. (2008a). Formalising dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200-15.

Marais, I., & Andrich, D. (2008b). Effects of Varying Magnitude and Patterns of Response Dependence in the Unidimensional Rasch Model. *Journal of Applied Measurement*, *9*(2), 1-20.

Masters G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

Mavhunga, E., & Rollnick, M. (2013). Improving PCK of Chemical Equilibrium in Pre-service Teachers. *African Journal of Research in Mathematics, Science and Technology Education, 17*(1-2), 113-125.

MedCalc Statistical Software version 15.11.0 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2015).

Park, S., & Chen, Y-C. (2012). Mapping Out the Integration of the Components of Pedagogical Content Knowledge (PCK): Examples From High School Biology Classrooms. *Journal of Research in Science Teaching, 49*(7), 922–941.

Park, S., Jang, J-Y., Chen, Y-C., & Jung, J. (2011). Is Pedagogical Content Knowledge (PCK) Necessary for Reformed Science Teaching?: Evidence from an Empirical Study. *Research in Science Education*, *41*, 245-260.

Pitjeng, P. (2014). Novice unqualified graduate science teachers' topic specific pedagogical content knowledge, content knowledge and their beliefs about teaching. In H. Venkat, M. Rollnick, M. Askew and J. Loughran (Eds). *Exploring Mathematics and Science Teachers' Knowledge: Windows into teacher thinking*, (pp. 65-83). Oxford: Routledge.

Rasch G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago Press. (The original work was published in Denmark in 1960.)

Rohaan, E. J., Taconis, R., & Jochems, W. M. G. (2011). Exploring the Underlying Components of Primary School Teachers' Pedagogical Content Knowledge for Technology Education. *Eurasia Journal of Mathematics, Science & Technology Education, 7*(4), 293-304.

Rollnick, M., Bennett, J., Rhemtula, M., Dharsey, N., & Ndlovu, T. (2008). The Place of Subject Matter Knowledge in Pedagogical Content Knowledge: A case study of South African teachers teaching the amount of substance and chemical equilibrium. *International Journal of Science Education*, *30*(10), 1365-1387.

Rollnick M., & Mavhunga, E. (2014). PCK of teaching electrochemistry in chemistry teachers: A case in Johannesburg, Gauteng Province, South Africa. *Educacíon Química, 25*(3), 354-362.

Schneider, R. M., & Plasman, K. (2011). Science Teacher Learning Progressions: A Review of Science Teachers' Pedagogical Content Knowledge Development. *Review of Educational Research, 81*(4), 530–565.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching, *Educational Researcher*, *15*, 4-14.

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

Smith, E. V. (2002). Understanding Rasch Measurement: Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement*, *3*(2), 205-31.

Smith, E. V. & Smith, R. M. (2004). *Introduction to Rasch measurement* (pp. 258-278). Maple Grove, MN: JAM Press.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research, 57*(8), 1358-1362.

Tepner, O. & Dollny, S. (2014). Measuring Chemistry Teachers' Content Knowledge: Is It Correlated to Pedagogical Content Knowledge? In C Bruguière, A. Tiberghien and P. Clément (Eds). Topics and Trends in Current Science Edu*cation: 9th ESERA Conference Selected Contributions,* Contributions from Science Education Research 1. (pp. 243-254). Springer Science+Business Media: Dordrect,

Van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing Science Teachers' Pedagogical Content Knowledge. *Journal of Research in Science Teaching*, *35*(6), 673-695.

Veal, W. R., & MaKinster, J. G. (1999). Pedagogical Content Knowledge Taxonomies. *Electronic Journal of Science Education, 3*(4). Retrieved from http://wolfweb.unr.edu/homepage/crowther/ejse/vealmak.html

Vokwana, N. Q. (2013). *Development and validation of instruments to assess content knowledge and topic specific pedagogical content knowledge of teachers of organic chemistry* (Masters dissertation). Retrieved from OpenUCT, http://hdl.handle.net/11427/6634

Wright, B., & Stone, M. (1999). *Measurement Essentials*. 2nd Edition. Wilmington, DE: Wide Range, Inc.