On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions...

Article · April 2017

DOI: 10.1016/j.compedu.2017.04.012

3 authors:


Ronny Scherer
University of Oslo


Jo Tondeur
Vrije Universiteit Brussel


Fazilat Siddiq
University of Oslo

On the Quest for Validity: Testing the Factor Structure and Measurement Invariance of the Technology-Dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) Model

Ronny Scherer

University of Oslo, Norway

Jo Tondeur

Vrije Universiteit Brussel, Belgium

Fazilat Siddiq

Nordic Institute for Studies in Innovation, Research and Education (NIFU), Norway

Author Note

Ronny Scherer, Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo; Jo Tondeur, Interfaculty Department for Teacher Education, Vrije Universiteit Brussel, Belgium; Fazilat Siddiq, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway.

Correspondence concerning this article should be addressed to Ronny Scherer, University of Oslo, Faculty of Educational Sciences, Centre for Educational Measurement at the University of Oslo (CEMO), Postbox 1161 Blindern, 0318 Oslo, Norway. E-Mail: ronny.scherer@cemo.uio.no

(Scherer, Tondeur, & Siddiq, 2017)

Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education, 112*, 1-17. doi:10.1016/j.compedu.2017.04.012

Highlights

- A general TPACK factor and a specific TK can be disentangled.
- The T-dimensions in the TPACK are highly correlated ($\rho > .80$).
- Measurement invariance across gender and educational track can be established.
- Significant gender differences in the T-dimensions exist.
- In conclusion, the TK factor stands out among all T-dimensions.

Abstract

The Technological, Pedagogical, and Content Knowledge (TPACK) framework – a framework which proposes a set of knowledge domains that are essential for effective teaching with technology – has gained considerable attention in the domain of education and technology. With the efforts to conceptualize these knowledge domains comes the question to what extent they can be distinguished empirically. Hence, the present study examines a measure that assesses pre-service teachers' self-efficacy in the technology-related TPACK dimensions ("T-dimensions"). In pursuit of crafting a validity argument, we investigated its factor structure and tested it for measurement invariance across gender and educational tracks, two subgroups that may indicate considerable differences. By means of multi-group confirmatory factor analysis, the data of $N = 665$ pre-service teachers in 18 teacher training institutions in Flanders (Belgium) revealed a nested factor structure of the TPACK measure, which comprised a general factor and a specific factor of pre-service teachers' technological knowledge. This factor structure was fully invariant across gender and educational tracks. Mean differences between educational tracks did not occur; yet, substantial differences were found across gender in favor of male pre-service teachers. This study sheds light on critical aspects of crafting a validity argument for the measurement of the T-dimensions in the TPACK framework and reports relevant subgroup differences.

Keywords: Gender differences; Measurement invariance; Teacher education; Technological, pedagogical, and content knowledge (TPACK); Validity

On the Quest for Validity: Testing the Factor Structure and Measurement Invariance of the Technology-Dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) Model

Introduction

Given the rapid development of information and communication technology (ICT) across almost all sectors in our society, educators face the challenge to help students develop competences that have become essential in the 21st century (Binkley et al., 2012). Of particular importance in this context are students' digital competences, which include, for instance, the abilities to collect, manage, produce, and exchange information in order to participate effectively in various sectors of society (Fraillon, Ainley, Schulz, Friedman, & Gebhardt, 2014). Teachers are therefore asked to foster the development of these competences and integrate ICT into their teaching (Graham, Borup, & Smith, 2012; Hew & Brush, 2006). Hence, the question arises to what extent teachers feel competent enough to accomplish this, and, moreover, which are the specific competences teachers should acquire?

As a response to the latter question, Mishra and Koehler (2006) proposed a conceptual framework of teachers' Technological, Pedagogical, and Content Knowledge (TPACK), clarifying which competences are needed for a successful integration of ICT into teaching and learning. Specifically, the TPACK framework distinguishes between general knowledge domains – content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) – and technology-dimensions ("T-dimensions") – technological content knowledge (TCK), technological pedagogical knowledge (TPK), technological pedagogical content knowledge (TPCK), and technological knowledge (TK) (Koehler, Mishra, Kereluik, Shin, & Graham, 2014). In order to draw inferences from this framework, especially with respect to potential consequences for teacher education and professional development, valid assessments of the TPACK knowledge domains are necessary (Koehler, Shin, & Mishra, 2012). On the quest for validity, researchers consider evidence on the factor structure, relations to cognate or distinct constructs, the comparability and therefore generalizability of measures across subgroups or populations to be essential in order to create a validity argument (AERA, APA, & NCME, 2014; Messick, 1995). Along these lines, existing studies attempted to validate TPACK self-report assessments by investigating their factor structure and the degree to which the knowledge domains can be empirically identified (Voogt, Fisser, Pareja Roblin, Tondeur, & van Braak, 2013). Nevertheless, the body of research abounds in mixed results, as indicated by the lack of replicability of the factor structure across samples of pre- and in-service teachers (e.g., Archambault & Barnett, 2010; Kopcha, Ottenbreit-Leftwich, Jung, & Baser, 2014). Moreover, to our knowledge, the comparability of self-report TPACK measures across relevant subgroups of teachers (e.g., gender and education programs) has not yet been addressed, although a sufficient degree of comparability needs to be established in order to conduct valid group comparisons. This is in fact surprising because crafting a reasonable validity argument requires tests of the extent to which a measure functions equally well across subgroups that are potentially subject to differences in the construct (AERA et al., 2014).

Clearly, TPACK seems very appealing to both researchers and teacher educators for describing the knowledge and skills that are needed for the effective ICT integration in teaching. In many studies, (pre-service) teachers' TPACK is measured through self-assessment surveys, but only a few studies provide a clear description of the instrument itself (Fisser, Voogt, Tondeur, & van Braak, 2015). Moreover, it seems that the boundaries between the TPACK knowledge domains are often fuzzy (Sang, Tondeur, Chai, & Dong, 2016). Against this background, the present study seeks to add to the existing literature by investigating the factor structure of a measure that represents pre-service teachers' *self-efficacy* in the T-dimensions of the TPACK framework. Further, it examines the extent to which the factor structure is invariant across gender and educational track, such that differences across the two groups can be investigated. The main contribution of this research lies in the discussion of aspects of the construct validity of TPACK measures.

Theoretical framework

The TPACK framework and its T-dimensions

The "Technological, Pedagogical, and Content Knowledge (TPACK)" model describes a framework consisting of different kinds of knowledge domains teachers need to become proficient in for successfully integrating digital technology in teaching and learning processes (Koehler et al., 2014). It was proposed by Mishra and Koehler (2006), and has become a dominant conceptual framework. Specifically, teachers and researchers have used this framework to (a) describe the competences pre- and in-service teachers should develop in order to deal with technology in 21st century education, and (b) understand and advance teachers' integration of digital technology in teaching and learning (Kopcha et al., 2014). Hence, the TPACK framework has consequently "(…) influenced theory, research, and practice in teacher education and teacher professional development" (Koehler et al., 2014, p. 101). TPACK is based on the notion of pedagogical content knowledge (PCK), which was introduced by Shulman (1986). This notion refers to the ability to combine content knowledge in a specific domain or school subject with pedagogical approaches to foster student learning (Voogt, Fisser, Pareja Roblin, et al., 2013). Shulman (1987) proposed three knowledge domains outside the context of technology (see also Mishra & Koehler, 2006; D. A. Schmidt et al., 2009):

- *Content knowledge (CK)* – domain-specific knowledge about the subject matter that teachers are supposed to teach.

- *Pedagogical knowledge (PK)* – knowledge about instructional practices, principles, and strategies to manage classrooms and organize the teaching of the subject matter.
- *Pedagogical content knowledge (PCK)* – knowledge about what instructional approaches fit the subject matter; it represents a synthesis of both content and pedagogical knowledge.

In this context, knowledge (K) refers to "a body of professional knowledge that encompasses both knowledge of general pedagogical principles and skills and knowledge of the subject matter to be taught" (Grossman & Richert, 1988, p. 54). Mishra and Koehler (2006) extended these three knowledge domains to the area of teaching and learning with technology and proposed a conceptualization of TCK, TPK, and TPCK. They further added teachers' knowledge about technology (TK). These "T-dimensions" are described as follows (e.g., Chai, Koh, & Tsai, 2013; Koehler et al., 2014; Koh, Chai, & Tsai, 2013; D. A. Schmidt et al., 2009):

- *Technological content knowledge (TCK)* – knowledge about how the subject matter can be represented with the help of technology; it includes knowledge about the reciprocal relation between content and technology, and the constraints of content knowledge by the capabilities of technology.
- *Technological pedagogical knowledge (TPK)* – knowledge about using technology to implement instructional practices, principles, and strategies; it includes an understanding that technology may change teaching.
- *Technological pedagogical content knowledge (TPCK)* – "knowledge about the complex relations among technology, pedagogy, and content that enable teachers to develop appropriate and context-specific teaching strategies"

(Koehler et al., 2014, p. 102); it forms the prerequisite for integrating technology into teaching.

- *Technological knowledge (TK)* – knowledge about various traditional and new technologies.

In the TPACK framework, TCK, TPK, and TPCK represent components that describe the interactions between Shulman's general knowledge domains and technology; TK appears to be a unique knowledge component that is comparable to teachers' content knowledge; yet, in the case of TK, the content refers to the technologies (D. A. Schmidt et al., 2009).

Although this framework proposes a set of related, yet distinct knowledge domains, there is only limited evidence on the empirical distinction between them. For instance, for the non-technical domains in the TPACK framework (i.e., CK, PK, and PCK), a number of studies found high correlations up to .96, particularly between CK and PCK (e.g., Krauss, Baumert, & Blum, 2008; Krauss, Brunner, et al., 2008; Paulick, Großschedl, Harms, & Möller, 2016). Nevertheless, these correlations varied across school types and teacher education programs (Kleickmann et al., 2013; Krauss et al., 2013). Studies that focused on the entire TPACK framework identified moderate to high correlations among all knowledge domains with a range between approximately .25 and .85 (e.g., Kaya & Dağ, 2013; Sahin, 2011). Interestingly, some patterns in these correlations can be observed: First, correlations *within* the non-technical or technical dimensions were usually higher than those *across* these dimensions. For example, Koh et al. (2013) estimated a correlation between TPCK and TPK of .74, and a correlation between TPCK and PCK was .23. A similar pattern was reported by D. A. Schmidt et al. (2009). Second, TK shows the lowest correlations to the other T-dimensions (e.g., Kaya & Dağ, 2013). Nevertheless, some of these observations were challenged by studies that revealed high correlations between the technical and non-technical dimensions, and low correlations among the T-dimensions (Chai, Koh, & Tsai, 2010; Kopcha et al., 2014).

The results on the empirical distinction between the TPACK dimensions appear to be rather mixed. Nevertheless, Koehler et al. (2014) argued that the "high degree of correlation between the subscales of TPACK raise questions about the extent to which the components of TPACK are, in fact, separate components" (p. 106). As a result, Chai, Koh, and Tsai (2016) concluded that "(…) when the factors are analyzed together, construct validity for all seven factors may be problematic" (p. 90), and Graham (2011) pointed out that the unclear boundaries between the TPACK knowledge domains still calls for further theoretical development and empirical research. Since the distinction between the TPACK dimensions seems to be unclear in situations where they are all assessed jointly, some researchers suggest conducting in-depth studies on either the technical or the non-technical dimensions in order to understand the nature of these dimensions (Voogt, Fisser, van Braak, & Tondeur, 2013). In response to this, the present study focuses on the T-dimensions in the TPACK framework.

Self-report measures of the T-dimensions

Current assessment practices of the T-dimensions within the TPACK framework comprise at least five types of assessments, including self-report measures, open-ended questionnaires, performance-based assessments, interviews, and observations in the classroom or during professional development (Chai et al., 2016; Koehler et al., 2014). In their review, Koehler et al. (2012) found that the majority of the TPACK assessments relied on pre- or in-service teachers' self-reports, which reflect their self-efficacy beliefs (Ay, Karadağ, & Acat, 2015; Koehler et al., 2014; Koh et al., 2013; Olofson, Swallow, & Neumann, 2016; D. A. Schmidt et al., 2009). In the context of social cognitive theory, these beliefs are defined as individuals' perceptions of their capabilities to plan and execute specific behaviour (Bandura, 1997), which affect a person's goals, actions, and effort (Skaalvik & Skaalvik, 2007). They can therefore be regarded as personal beliefs about what a person can do (Bong & Skaalvik, 2003; Scherer, Jansen, Nilsen, Areepattamannil, & Marsh, 2016).With respect to the T-dimensions, teachers are asked to indicate the degree to which they feel confident to perform a number of tasks that are directly related to TK (e.g., I know how to solve my own technical problems), TCK (e.g., I know about technologies that I can use for understanding and doing literacy), TPK (e.g., I can choose technologies that enhance the teaching approaches for a lesson), and TPCK (e.g., I can teach lessons that appropriately combine literacy, technologies, and teaching approaches; D. A. Schmidt et al., 2009).

Although self-efficacy measures of the T-dimensions do not represent performance-based indicators, yet respondents' self-perceptions, they have been widely used in educational technology research (Abbitt, 2011; Chai, Koh, Tsai, & Tan, 2011; Jen, Yeh, Hsu, Wu, & Chen, 2016; Koh, Chai, & Tay, 2014; Kopcha et al., 2014). We do see a number of reasons for this choice of measures: First, self-efficacy measures are easily accessible and provide a cost-efficient measurement of competence perceptions; in fact, they provide reliable and valid indicators of one's self-beliefs (Bandura, 1997; Banoglu, Vanderlinde, & Yildiz, 2015; Scherer & Siddiq, 2015). Second, self-efficacy beliefs are important determinants for teachers' intention to use and integrate technology for teaching and learning (Scherer, Siddiq, & Teo, 2015; Teo, 2011). Third, teachers' perceptions of their competences with respect to TPACK provide information about their beliefs about how well they will be able to perform these competences in future situations (Bandura, 1997). Hence, they are forward-oriented and determine the future integration of ICT, teachers' motivation for competence development, and their job satisfaction (Chesnut & Burley, 2015; OECD, 2014; Scherer, Jansen, et al., 2016). Fourth, teachers' self-efficacy beliefs correlate with the quality of their instruction and, in turn, with students' educational outcomes, such as achievement, interest, and motivation (Holzberger, Philipp, & Kunter, 2013; OECD, 2014; Zee, de Jong, & Koomen, 2016). In light of this relation, self-efficacy measures are often regarded as proxies of competences, although their correspondence is not perfect. Voogt, Fisser, Pareja Roblin, et al. (2013) consequently argued that teacher competences – often measured as their proficiency in knowledge domains – and beliefs are intertwined (p. 109). Fifth, self-efficacy beliefs are integral parts of teachers' belief system (Antonietti & Giorgetti, 2006; Klassen & Tze, 2014), and are therefore considered to be important outcomes of teacher education and professional development (Chesnut & Burley, 2015; OECD, 2014). Against this backdrop, the importance of teachers' self-efficacy beliefs can clearly be established.

TPACK measures across subgroups: Gender and educational tracks

In order to interpret self-report TPACK measures appropriately, a number of constraints need to be considered: First, as these measures represent a person's beliefs about his or her competence, they are subject to *individual differences* (Bandura, 1997; Jansen, Scherer, & Schroeders, 2015). Second, Tschannen-Moran and Hoy (2007) pointed out that mastery experiences are the main sources of self-efficacy beliefs, thereby stressing the fact that *contextual differences* in the resultant self-efficacy measures may occur (see also Tschannen-Moran & Johnson, 2011). Put differently, systematic differences in teachers' self-efficacy beliefs may be due to systematic differences in the school or teacher training environments, and can lead to differences between subgroups of teachers.

Along these lines, existing research has examined potential gender gaps in the context of ICT (Scherer & Siddiq, 2015; Tondeur, Van de Velde, Vermeersch, & Van Houtte, 2016); as a matter of fact, knowledge about the gaps in self-efficacy beliefs has been considered crucial, because it may uncover differences in teachers' intentions and actions to use ICT for teaching and learning (e.g., Baek, Jung, & Kim, 2008; Plumm, 2008; Tondeur, van Keer, van Braak, & Valcke, 2008). Currently, gender differences in teachers' self-efficacy beliefs are controversially discussed, and the existing body of research abounds in mixed results (Sang, Valcke, Braak, & Tondeur, 2010; Scherer & Siddiq, 2015). For example, a number of studies found that females tend to consider themselves as less competent in using computers than males, although they may be digitally competent to the same extent (Durndell & Haag, 2002; Scherer & Siddiq, 2015; Sieverding & Koch, 2009). Taking a motivational perspective, Koch, Müller, and Sieverding (2008) argued that the attributions to failure may also play an important explanatory role; specifically, whereas females attributed their failure in solving computer-related tasks to themselves, males attributed their failure to external factors. Although these studies leave us with reasons to believe that gender differences in self-efficacy beliefs may exist in the context of ICT, a number of studies failed to identify these differences (e.g., Compton, Burkett, & Burkett, 2003; Pamuk & Peker, 2009). In light of these findings, we conclude that it is important to take a gender perspective on the T-dimensions in order to disentangle potential differences or even gaps.

As mentioned earlier, contextual effects may also determine teachers' self-efficacy beliefs about the T-dimensions. This hypothesis is grounded in the idea that mastery experiences and the opportunities to learn that are offered within learning environments – be it during teaching in schools or during professional development and teacher education in teacher training institutions – are potential sources of self-efficacy (Blömeke, Suhl, Kaiser, & Döhrmann, 2012; W. H. Schmidt, Cogan, & Houang, 2011; Tschannen-Moran & Hoy, 2007). Put differently, variation in teachers' TPACK beliefs may be related to the variation in the opportunities of mastery experience they are provided with in their specific working or educational context. In fact, there is considerable evidence that these opportunities vary across the different educational tracks pre-service teachers are enrolled in (Tatto et al., 2012). Moreover, Krauss et al. (2013) have shown that considerable differences in teachers' content knowledge in the domain of mathematics existed between teachers in academic and non-academic tracks. They argued that teacher education differed across these tracks, particularly in the level and amount of content knowledge courses. Nevertheless, the same study did not reveal any differences in pedagogical content knowledge[1]. In light of these results, we consider it important to examine whether differences in teachers' TPACK exist across different educational tracks, as they may help us identify the potential for professional development and interventions to strengthen teachers' TPACK. Furthermore, knowledge about these differences extends the existing body of literature and increases our understanding about the nature of TPACK in educational contexts.

Along with the investigation of gender and educational track differences comes the question to what extent the TPACK self-efficacy measures can be generalized across these subgroups. This question is of particular importance, because it addresses the construct validity of measures (Messick, 1995). In fact, if the TPACK measures work differently across gender and educational tracks, group comparisons are not valid, as mean differences may only be due to the differential functioning of the TPACK items (Millsap, 2011). Since this would compromise the validity of inferences across subgroups, especially with respect to mean comparisons and comparisons of regression coefficients (Guenole & Brown, 2014), it is necessary to test the invariance of the TPACK measures. Although the examination of measurement invariance has gained attention in the domain of educational technology (e.g., Scherer & Siddiq, 2015; Teo, Lee, Chai, & Wong, 2009), we are not aware of any study that inquired different levels of invariance for existing TPACK measures.

The present study

TPACK has been adopted by many researchers and practitioners for describing the knowledge and skills that are needed for the effective ICT integration in education (Graham, 2011; Koh, Chai, & Tsai, 2010). Nevertheless, TPACK surveys are still in the process of construct validation, and a reliable and validated instrument for measuring pre-service

[1] Note that both knowledge domains (i.e., CK and PCK) were measured by performance-based tests in this study.

teachers' TPACK is still lacking (Sang et al., 2016). To illustrate, the knowledge domains of the TPACK framework could not be reproduced through exploratory factor analysis (Archambault & Barnett, 2010). Considering the construct validation challenges, the present study's purpose is to contribute to crafting a validity argument for a tool that empirically measures and describes the technology-dimensions within TPACK.

Specifically, focusing on the T-dimensions, the present study seeks to test the factor structure of the corresponding measure for a sample of pre-service teachers. Given the conceptualization of the T-dimensions in the TPACK framework, we expect them to be related but still distinct. This expectation would manifest in moderate to high factor correlations below 1. Because the measurement of the T-dimensions might be subject to considerable differences across gender and the educational track pre-service teachers participate in, we investigate the measurement invariance and the significance of mean differences between these groups. Until now, measurement invariance has largely been ignored in the context of measuring TPACK. This is quite surprising given that evidence on invariance across relevant subgroups is considered to be critical to the creation of a validity argument. If measurement invariance can indeed be established to a sufficient degree, mean comparisons between the gender and educational track groups can be employed to inform researchers and teacher educators about potential deficits in pre-service teachers' perceived TPACK. The corresponding research questions read:

1. *To what extent can the T-dimensions (i.e., TCK, TPK, TPCK, and TK) be distinguished in the TPACK measure? (Factor structure)*
2. *To what extent does the measurement model of the T-dimensions show invariance across pre-service teachers' gender and educational track? (Measurement invariance)*

Method

Sample and procedure[2]

The present study is based on a sample of *N* = 665 last-year pre-service teachers in eighteen teacher training institutions who participated in an online survey in 2014 in Flanders, the Dutch-speaking part of Belgium. At the outset of the study, we contacted the department heads of twenty-one teacher training institutions in Flanders, twenty of which were willing to contribute to our study by allowing us to administer the online questionnaire to their teacher training classes (see also Authors, 2016). After processing the resultant data from the convenience sample of 688 pre-service teachers, twenty-three cases had to be excluded due to either extreme responses (i.e., pre-service teachers only used the lowest or highest response categories) or missing data on all relevant variables. Participation in this survey was anonymous. Pre-service teachers responded to the items assessing the T-dimensions of the TPACK framework, their background (e.g., age, gender, educational track), and further variables related to the use and intentions to integrate digital technologies in their teaching. In total, 73.8% of the participants were females; in fact, this gender distribution is representative of the pre-service teachers in Flanders (Tondeur, van Braak, Siddiq, & Scherer, 2016). The average age was 25.1 years (*SD* = 7.7 years). Of these pre-service teachers, 57.7% had obtained a Bachelor's degree in higher education, whereas 42.3% had obtained a specific teacher training degree from universities, colleges, or centers for adult learning across various subjects that ranged from arts education to physical education.

Measures

Measurement of the T-dimensions. The measurement of the T-dimensions (TCK, TPK, TPCK and TK) was based on the adapted Dutch version of Schmidt et al.'s (2009)

[2] Please note that the data used in the present study have been used in a previous publication on validating a measure of pre-service teachers' perceptions of effective training strategies to foster the integration of educational technology (Authors, 2016). Hence, some parts of the method section were adapted from this publication in which the same data were used.

TPACK self-report scale (Fisser, Voogt, Van Braak, & Tondeur, 2013). Teachers were asked to indicate their agreement with a number of statements that referred to these dimensions on a five-point scale (from *0 = I completely disagree* to *4 = I completely agree*), such that lower values correspond to low self-efficacy and higher values to high self-efficacy in the corresponding T-dimensions. The corresponding item wordings and descriptive statistics are detailed in Table 1.[3]

Educational track. Pre-service teachers were asked about their educational track, that is, the type of teacher training they followed. Specifically, they could choose among the following five options: Bachelor's degree in primary education, Bachelor's degree in secondary education, Special teacher training at University, College, or Adult Education Centers. The first two options were merged to the category "Bachelor's degree in education", because they basically led to the same type of degree pre-service teachers would gain after completion of the program, and they contained similar education modules; the remaining categories were merged to "Special teacher training". This resulted in a dichotomized indicator of pre-service teachers' educational track (*0 = Specific teacher training*, *1 = Bachelor's degree in education*). The three-year training program for primary and lower secondary teachers is offered as a professional bachelor training program at colleges of higher education; the specific teacher training is designed for students who have already obtained a diploma in higher or adult education, and is provided by universities, adult education centers, and also by colleges of higher education.

Data analysis

Measurement models and estimator. In order to evaluate the goodness-of-fit of structural equation models specified in the present study, we referred to common guidelines for an acceptable model fit (i.e., CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .08, and SRMR ≤ .10;

[3] Please find further summary statistics in Table B1 in the Appendix.

Marsh, Hau, & Grayson, 2005). In all analyses, robust maximum likelihood (MLR) estimation with standard errors and tests of fit that were robust against non-normality of observations and the use of categorical variables in the presence of at least four response categories was used (Rhemtulla, Brosseau-Liard, & Savalei, 2012). The MLR continuous estimation can handle missing values that are missing at random more appropriately than, for instance, the categorical weighted least squares means and variance adjusted (WLSMV) estimation (Asparouhov & Muthén, 2010; Beauducel & Herzberg, 2006). As a bifactor model distinguishing between a general TPACK factor and four specific factors represents the most complex measurement model with 84 free parameters to be estimated in this study, we considered the minimal sample size : number of estimated parameters ratio of 665:84 = 7.9:1 to be sufficiently large in order to detect the proposed number of factors (Kline, 2005). Moreover, this ratio falls into the recommended guidelines to treat item responses continuously (Lei & Wu, 2012). For more parsimonious measurement models, this ratio is larger.

Reliability. As measures of internal consistency, we estimated Cronbach's α and McDonald's ω[4]. The latter is based on unconstrained assumptions on the measurement models of factors (i.e., freely estimated factor loadings and residual variances; McDonald, 1999). In fact, if items within a scale are heterogeneous, McDonald's ω provides a more precise reliability estimate than Cronbach's α (Revelle & Zinbarg, 2008). For a single-factor model comprising K items with standardized factor loadings λ_i and standardized residual variances δ_{ii}, and based on the assumption that residual correlations do not exist, McDonald's ω is calculated as $\omega = [(\sum_{i=1}^{K} \lambda_i)^2]/[(\sum_{i=1}^{K} \lambda_i)^2 + \sum_{i=1}^{K} \delta_{ii}]$ (Gignac, 2009). All scales showed acceptable reliabilities (Table 1); these reliabilities fell within the boundaries a recent

[4] McDonald's ω was specified as ω_h, that is, ω-hierarchical (Revelle & Zinbarg, 2008).

systematic review of TPACK measures has identified (i.e., internal consistencies >.70; Voogt, Fisser, Pareja Roblin, et al., 2013).

Measurement invariance testing. As outlined earlier, it has been unclear whether the existing measures of the T-dimensions are comparable across gender and educational tracks. In order to address this gap, we conducted measurement invariance testing by means of multi-group confirmatory factor analysis (MG-CFA; Research Question 2). Specifically, we tested the TPACK measurement model for configural, metric, and scalar invariance by systematically constraining the factor loadings and item intercepts to equality across gender and educational tracks. In this respect, the *configural* invariance model assumes the same number of factors (i.e., latent variables) and item-factor links (i.e., items load on the same factors in each group) across groups; yet, all model parameters (i.e., factor loadings, item intercepts, and residual variances) are freely estimated for each group. The next level of invariance is referred to as *metric* invariance. In a MG-CFA model assuming metric invariance, the factor loadings are constrained to equality across groups, putting the latent variables on the same scale. Imposing equality constraints on the item intercepts (i.e., the means of item responses) leads to the *scalar* invariance model. If this level of invariance is achieved, comparisons of factor means across groups are justified (Millsap, 2011).

In order to decide which level of invariance can be achieved, we evaluated each model with respect to its goodness-of-fit indices on the one hand, and compared the differences in goodness-of-fit statistics between the configural model (baseline) and the further levels of invariance on the other hand (Marsh et al., 2009). In the psychometric literature, a number of cut-off values for changes in the CFI, RMSEA, and SRMR have been suggested. For instance, Chen (2007) proposed considering values of ΔCFI $\leq$ -.010, ΔRMSEA $\leq$ .015, and ΔSRMR $\leq$ .030 to be indicative of insignificant changes in model fit when comparing the more restrictive with the less restrictive invariance model. Cheung and Rensvold (2002) proposed similar cut-off values for the CFI (ΔCFI ≤ -.010), whereas Meade, Johnson, and Braddy (2008) provided stricter values, ΔCFI ≤ -.002. Finally, Khojasteh and Lo (2015) indicated more conservative criteria for the RMSEA in situations, where the measurement model follows a bi-factor structure, ΔCFI ≤ -0.04, ΔRMSEA ≤ .034, ΔSRMR ≤ .030. In addition, χ^2 difference testing can be employed (Brown, 2015). At this point, it must be noted that these cut-off values cannot be regarded as “golden rules” (Marsh, Hau, & Wen, 2004). Instead, they provide orientations toward the degree of substantial model fit differences. Furthermore, their performance varies with respect to sample size, the number of latent variables, the treatment of the data (continuous vs. categorical), the number of groups, the type of measurement invariance tested, and the factor structure specified (Khojasteh & Lo, 2015; Meade et al., 2008; Rutkowski & Svetina, 2014). In the present study, we followed the cut-off-values Khojasteh and Lo (2015) suggested, because they enable us to evaluate the invariance of complex measurement models (e.g., nested factor models).

Handling missing data and the clustered sample structure. In the current study, a relatively small amount of missing data in the response variables occurred (up to 3.3%). Since these missing values were not due to the design of the study, we assumed that they occurred randomly. As a consequence, we applied the full-information maximum likelihood estimation, which handles the occurrence of missing data (Enders, 2010). Notice that we did not decide on applying multiple imputation to the dataset for two main reasons: First, full-information maximum likelihood estimation and multiple imputation perform almost equally in situations where missing data are at random and for small rates of missingness (e.g., Enders, 2010). Second, the performance of multiple imputation results in a number of complete datasets, for which the proposed structural equation models are fitted separately; the resultant model parameters are pooled in a final step. To this end, only little is known about, for instance, the performance of fit indices in measurement invariance testing situations for pooled models (Enders & Mansolf, 2016).

Given that student teachers were enrolled in one of 18 teacher education institutions, the present data follow a clustered sample structure (i.e., student teachers clustered in institutions). Because neglecting that student teachers within an institution might be more homogeneous in their responses on the TPACK items than between institutions may lead to substantial bias in the estimation of structural parameters such as regression coefficients or correlations (Snijders & Bosker, 2012), we accounted for the clustered data structure by adjusting the χ^2 statistic and the standard errors of all model parameters in the statistical package M*plus* 7.3 (TYPE = COMPLEX option; Muthén & Muthén, 1998-2015). Hence, the χ^2 statistics and the results of χ^2 difference testing are reported after applying the Satorra-Bentler correction (Satorra & Bentler, 2010).

Results

Factor structure (RQ1)

Research Question 1 was concerned with the factor structure of the T-dimensions measure. We hypothesized that the four T-dimensions can be disentangled empirically, assuming that a confirmatory factor-analytic model with four correlated traits represented the factor structure. To test this assumption, we conducted a number of analytical steps: First, we specified a unidimensional model, which represented the T-dimensions by a single factor (Figure 2a). This model showed a marginal model fit and was therefore rejected (Table 2; Single-factor model). Second, we specified a model that assumed four correlated factors (i.e., TCK, TPK, TPCK, and TK; Figure 2b). The resultant model showed a good fit and consequently pointed to the benefit of distinguishing between the four T-dimensions (Table 2; Four-factor model). Nevertheless, the correlations among the four factors were rather high (see Table 3), particularly between the dimensions of TCK, TPK, and TPCK (ρ = .98–.99);

lower correlations were observed for TK (ρ = .81–.86). These findings suggested high dependencies among pre-service teachers' self-beliefs on the T-dimensions and implied the existence of a general factor that underlies the four dimensions; yet, the TK factor seems to stand out in this respect, as the lower correlations indicated. Hence, in a third step, we specified a two-factor model that distinguished between the TK factor and a factor that combined the other dimensions (Figure 2c). The resultant model fitted the data significantly better than the unidimensional model (Table 2; Two-factor model) and revealed a high correlation between TK and the conglomerate factor, ρ = .83. This model was however not superior to the four-factor model; yet, the differences in model fit were rather low, *SB*-$\Delta\chi^2$ (5) = 18.1, $p < .01$. Although the correlation between TK and the combination of TCK, TPK, and TPCK was significantly lower than 1, which was indicated by the fact that the two-factor model outperformed the unidimensional model in terms of goodness-of-fit, the two factors still shared considerable variance. Following the approach proposed by Chen, West, and Sousa (2006) which suggests the use of bifactor or higher-order factor models if factors are highly correlated, we specified a bifactor model that comprised a general factor and four specific, uncorrelated factors in order to disentangle to what extent the four dimensions measure specific aspects after controlling for pre-service teachers' general self-efficacy beliefs in the T-dimensions (Figure 2d). Indeed, this bifactor model fitted the data well, and the assumption of a general factor improved the model fit significantly (Table 2; Bifactor model). However, none of the factor loadings of the TCK and TPCK factors were statistically significant. Moreover, only one factor loading was positive and significant for the TPK factor (Item *TPK 1*, λ = .15). In contrast, all TK factor loadings were positive and statistically significant, and ranged between .14 (Item *TK7*) and .56 (Item *TK 1*). These numbers indicated positive specificities of the TK items. In light of these results, we concluded that only the TK factor could be modeled as a specific factor beyond the general factor. The fourth and final step consequently resulted in a nested factor model, which only contained these two factors (Figure 3). Indeed, this model showed a good fit and outperformed the four-factor model significantly (Table 2; Nested factor model). Table 4 details the corresponding factor loadings and specificities[5]. Factor loadings of the general TPACK factor (λ_{GEN}) ranged between .53 and .78; factor loadings of the specific TK factor were lower, λ_{TK} = .13–.56. These loadings resulted in TK specificities between .04 (Item *TK7*) and .53 (Item *TK1*), and indicated that the TK dimension can be distinguished from the general factor. The nested factor model was accepted as the final measurement model for the total sample. We note that, with the exception of the bifactor modeling approach, two residual correlations were included in the measurement models (i.e., *TPK4—TPK5*, *TPCK1—TPCK5*). These correlations were specified on the basis of overlapping item formulations and modification indices[6]. Considering the information criteria across the five models tested in this study, the bifactor model showed the lowest AIC and sample-size adjusted BIC values, whereas the nested factor model showed the lowest BIC value (see Table 2). In comparison to the drops in these information criteria when specifying more factors in the TPACK measurement model, the differences between the bifactor and nested factor model were only marginal. Taking into account that only the specific factor representing TK showed both significant variation beyond that of the general TPACK factor and significant factor loadings, we argue that the nested factor model is "practically more relevant" than the bifactor model; the latter produces higher model complexity and more factors that could not be identified empirically (Brown, 2015; Kline, 2005). As Eid, Geiser, Koch, and Heene (2016) argue, in situations where the

[5] In this respect, specificity refers to the degree to which individual differences in the item responses can be explained by the specific TK factor over and above the variance explanation by the general TPACK factor (Rodriguez, Reise, & Haviland, 2015).

[6] In the software package M*plus*, modification indices were estimated in order to identify potential model fit improvements that may occur after the introduction of, for instance, residual correlations or cross-loadings (Chou & Huh, 2012).

bifactor model contains "irrelevant" specific factors, the model without these factors should be accepted.

In order to support the decision for the nested factor model with further empirical evidence, we employed exploratory structural equation modeling with up to four factors. This modeling approach was based on the "Geomin" rotation, which represents an oblique factor rotation method that is usually recommended if the factors are assumed to be correlated (for more details on the advantages of this rotation method, please refer to Marsh, Morin, Parker, & Kaur, 2014; Price, 2017). Tables A1 to A3 in the Appendix detail the factor loadings for the models with two, three, and four factors, and clearly show that substantial cross-loadings between all factors existed. These cross-loadings strengthen the assumption of a general TPACK factor.

In sum, our response to Research Question 1 is: The factor structure of the T-dimensions measure can be described by a general factor and a specific TK factor; a clear empirical distinction between the four T-dimensions could not be retained.

Measurement invariance across gender and educational track (RQ2)

Research Question 2 was concerned with the invariance of the nested factor model we identified under RQ1. This question fed into the discussion on the comparability of the factor structure that represents the T-dimensions across gender and educational tracks. Table 5 shows the resulting goodness-of-fit statistics and the corresponding model comparisons.

For both grouping variables, gender and educational track, the configural models fitted the data well. This provides evidence that the nested factor model can be applied to both gender and educational track groups. Furthermore, the metric and scalar invariance models showed good fit statistics. Regarding the changes in model fit, all criteria for the ΔCFI, ΔRMSEA, and the ΔSRMR values were met. Even χ^2 difference testing suggested the preference of the models with higher levels of invariance, except for the difference between the configural and scalar model across gender. We therefore concluded that scalar invariance held across both gender and educational tracks. Notice, we did not take into information criteria when comparing the measurement invariance models with each other. To our best knowledge, it is still unclear to what extent changes in information criteria are sensitive toward equality constraints in model parameters across groups, and whether specific thresholds for their differences can be derived. In accordance with Scherer, Nilsen, and Jansen (2016), we only took into account changes in the CFI, RMSEA, SRMR, and the χ^2 values – fit indices that are better understood in invariance testing (e.g., Cheung & Rensvold, 2002; Millsap, 2011; Rutkowski & Svetina, 2014). This result has at least two consequences: First, the nested factor structure of the T-dimensions measure applies to both gender and educational track groups. Second, mean comparisons of the general factor and the specific TK factor can be conducted.

The factor means with respect to pre-service teachers' gender were: Male students, $M_{TK} = 0.00$, $SD_{TK} = 0.43$, $M_{TPACK} = 0.00$, $SD_{TPACK} = 0.56$; Female students, $M_{TK} = –0.30$, $SD_{TK} = 0.59$, $M_{TPACK} = –0.17$, $SD_{TPACK} = 0.60$. Gender differences were statistically significant for both the general factor ($\Delta M_{TPACK} = 0.17$; $t = 3.27$, $p < .01$; $d = 0.29$, 95% CI $d = [0.11, 0.46]$) and the TK factor ($\Delta M_{TK} = 0.30$; $t = 6.15$, $p < .01$; $d = 0.54$, 95% CI $d = [0.37, 0.72]$) with moderate effect sizes. Hence, gender differences in the identified general TPACK and the specific TK factor existed in favor of male pre-service teachers.

Based on the scalar invariance model, the following unstandardized factor means across pre-service teachers' educational tracks for the general TPACK factor and the specific TK factor: Students enrolled in special teacher education programs, $M_{TK} = 0.00$, $SD_{TK} = 0.54$, $M_{TPACK} = 0.00$, $SD_{TPACK} = 0.62$; Students enrolled in a Bachelor program, $M_{TK} = –0.08$, $SD_{TK} = 0.62$, $M_{TPACK} = 0.08$, $SD_{TPACK} = 0.47$. Neither the mean difference in the general factor ($\Delta M_{TPACK} = –0.08$; $t = 1.89$, $p = .06$; $d = –0.15$, 95% CI $d = [–0.30, 0.06]$), nor the mean difference in the specific TK factor (ΔM_{TK} = 0.08; t = 1.73, p = .08; d = 0.14, 95% CI d = [–0.02, 0.29]) were statistically significant. Hence, there was no evidence on factor mean differences across educational tracks. We further note that interaction effects between gender and educational track have not been tested, as these would have been based on small sample sizes and therefore lose out on power to detect the proposed factorial structures.

Discussion

Factor structure and comparability of the technology-dimensions

Our findings on the factor structure of the T-dimensions measure revealed a general factor and a specific TK factor. Given that no further differentiation between the T-dimensions in the TPACK framework could be achieved in this study, the expectation that the T-dimensions provide a set of empirically distinct dimensions has not been met, and the underlying theoretical framework could only be partly confirmed. Interestingly, pre-service teachers' self-efficacy beliefs in their technological knowledge (TK) played a specific role among the T-dimensions, because it showed lower correlations to the other T-dimensions (TCK, TPK, and TPCK) and substantial specificity. This specificity may be due to the demands and conceptualization of TK – in contrast to TCK, TPCK, and TPK, TK does not include specific knowledge about how to teach with digital technology, thereby focusing on the technological aspect only (D. A. Schmidt et al., 2009). Pre-service teachers' self-efficacy beliefs in TK may therefore be distinct from their self-efficacy beliefs in the other dimensions. From a measurement point of view, one may argue that this finding points to at least three possible ways of handling scale scores: (a) report the single-factor TPACK scale score only; (b) consider the TK scale as completely distinct from the other scales; (c) delete the TK items from the TPACK scale. We believe that neither of these ways takes into account the complex factor structure of the TPACK T-dimensions appropriately. Option (a) neglects the fact that a single-factor model does not fit the data well and that the TK scale shows substantial specific variation beyond what can be explained by a general TPACK factor. Option (b) might in fact be the closest approximation to the data – as indicated by an improvement in model fit when specifying the TK factor next to another factor; yet, the high correlation between the two resultant factors does not provide sufficient evidence for discriminant validity, thus compromising the practical importance of two distinct factors. Option (c) disregards the fact that the TK items show psychometrically sound properties (e.g., in terms of reliability) and capture some specific variation beyond general TPACK. Hence, the picture of the TPACK factor structure is by no means clear. In fact, one needs to consider what the different factors in the nested TPACK model represent (Eid et al., 2016): The general TPACK factor captures variation in all item responses; in other words, it captures what is common to all T-dimension items. The specific TK factor captures the variation in the TK items after controlling for the general factor. This variation might represent a context effect, as it refers to pure technological knowledge independent of teaching and learning. Scherer and Siddiq (2015) argued similarly and showed that teachers' computer self-efficacy can indeed be differentiated into operational skills (i.e., the technological aspect) and instructional skills (i.e., the pedagogical aspect) in using digital technology. Furthermore, from a substantive point of view, the T-dimensions in the TPACK framework may naturally "cling together" for the sample of pre-service teachers, as indicated by the high factor correlations, which are higher than those reported for other studies and teacher samples (Chai et al., 2016; Kaya & Dağ, 2013; Koh et al., 2013). We suspect that the degree to which the T-dimensions can be differentiated may differ between pre- or in-service teacher samples. More specifically, there has been some evidence that teachers' self-efficacy differentiates over time and in the course of gaining working experience in school (Klassen & Chiu, 2010; Tschannen-Moran & Hoy, 2007). Hence, the factor structure and the distinction between the T-dimensions may be subject to change with teachers' experience. In our opinion, this calls for in-depth studies that compare TPACK measures across multiple samples of pre- and in-service teachers. Finally, we would like to point out that the high correlations among the T-dimensions have also been found in the non-technical domains. For instance, Baumert et al. (2010) identified a similarly high correlation between CK and PCK, which was moderated by the type of schools. It therefore seems as if these dimensions are generally closely related (Fisser, Voogt, Tondeur, & van Braak, 2015). In light of our findings on the factor structure of the T-dimensions measure (Research Question 1), we conclude that evidence on the validity of the measure could only be partly obtained. On the one hand, TK is specific within the T-dimensions; on the other hand, the relations among all T-dimensions are substantially high. We therefore encourage the further development of TPACK measures that might distinguish more clearly between the different knowledge dimensions and further validation studies that investigate the substantial meaning of the specific TK factor.

The present study adopted a correlated traits approach, thereby assuming the TPACK factor structure can be represented by a number of correlated latent variables. As a potential alternative, structural relations among the T-dimensions may be assumed (Chai et al., 2011; Koh et al., 2013). As Koh et al. (2013) suggested, indirect effects, TK → TPK → TPCK and TK → TCK → TPCK, describe the idea that technological knowledge forms the prerequisite for the pedagogical and content knowledge-related T-dimensions, which in turn predict TPCK. In fact, together with Chai et al. (2011), they were able to gather some evidence for these structural relations. This relation has also been hypothesized in non-technical domains. Baumert et al. (2010) found evidence for the hypothesis that CK forms the prerequisite for PCK in the domain of mathematics. In the present study, however, such structural relations were not adopted, as the analysis of the factor structure revealed only the existence of a general TPACK factor and a specific TK factor. Moreover, in light of the high correlations among the T-dimensions, the question arises to what extent single knowledge components will have a unique contribution to explaining variance in TPCK after controlling for the remaining components – this is also a methodological question that concerns the issue of multicollinearity in indirect effects models with highly correlated variables (e.g., Shrout & Bolger, 2002). Hence, it may be further investigated whether or not structural relations among the T-dimensions exist.

Regarding our second research question, we found support for the invariance of the TPACK measure across gender and educational tracks. This finding suggests that the factor structure – as identified in Research Question 1 – was replicable in these two subgroups, therefore providing evidence on the comparability of the proposed structure among subgroups. Moreover, the fact that even strong (scalar) invariance held, speaks for the comparability of the T-dimensions assessment and acceptable psychometric properties thereof. In fact, given that sufficient levels of invariance were achieved, mean differences in both the general TPACK and the specific TK factors can be interpreted as actual mean differences, which are not due to the differential functioning of the self-efficacy items (Millsap, 2011).

On the basis of the scalar invariance models, we examined factor mean differences. Although differences in the general TPACK factor and the specific TK factor did not occur between educational tracks, gender differences were identified. As mentioned earlier, these differences cannot be attributed to the differential functioning of the TPACK items, because measurement invariance was achieved; hence, they reflect actual mean differences in the latent variables. Had invariance not been met, the validity of the inferences made based on these differences would have been severely threatened (Zumbo, 2006).

Gender differences. The gender differences in pre-service teachers' self-efficacy with regard to the T-dimensions showed that males perceived themselves as more competent than females. This finding was somehow expected, because existing research clearly identified a "gender gap" in technology-related self-efficacy in favor of males (e.g., Scherer & Siddiq, 2015; Sieverding & Koch, 2009; Vekiri & Chronaki, 2008). Since our study has replicated this finding, it seems as if it is (still) a rather persistent observation. At the same time, it is important to note that gender did *not* serve as a moderating variable of the relation between the unobserved latent variables and the observed manifest items, which could have influenced the factor structure of the T-dimensions measure; yet, gender *differences in the factor means* existed. At this point, we can only speculate why these differences occurred. Potential factors that may have determined pre-service teachers' responses to the scales assessing the T-dimensions may relate to the tendency of females to underestimate their knowledge and skills (Sieverding & Koch, 2009) or gender-stereotypical beliefs (Koch et al., 2008). Despite the lack of explanatory factors in the present study, the gender differences still point to the need to strengthen self-beliefs in teaching with and handling technology in pre-service teacher education.

Educational track differences. Interestingly, educational track differences did not occur. This finding was somehow surprising, because differences in the opportunities to learn and therefore the opportunities to gain mastery experiences provide sources for self-efficacy beliefs – at least for students (Usher & Pajares, 2008) and in-service teachers (Tschannen-Moran & Hoy, 2007; Tschannen-Moran, Hoy, & Hoy, 1998). One potential explanation for this observation may lie in the fact that the two teacher training programs may not necessarily provide entirely different opportunities to gain mastery experiences in using digital technologies for teaching and learning purposes. Specifically, the extent to which the T-dimensions are emphasized could be comparable across tracks. Nevertheless, Baumert et al. (2010) pointed out that even mere structural differences in teacher education programs may already lead to differences in teacher knowledge domains. In addition, one should not forget that differences in the opportunities to gain mastery experiences are not the only factors that come into play in the formation of self-efficacy; Klassen and Tze (2014), for instance, argued that teachers' personality is another determining factor. In light of these considerations, we encourage further, in-depth studies on the sources for potential educational track differences in the TPACK dimensions, which take into account the specific curricula and opportunities to learn in the teacher training programs.

Limitations and future directions

The current study has some limitations that point to future directions for research on the measurement of TPACK: First, since the T-dimensions in the TPACK framework were measured by self-report items it is to be determined to what extent these self-reports are vulnerable and sensitive to response bias, which may manifest in specific response styles that are due to social desirability, overrating, or acquiescence (He, Bartram, Inceoglu, & van de Vijver, 2014). We therefore encourage researchers and test developers to examine the extent to which the TPACK self-efficacy measure corresponds to an actual performance-based measure of the T-dimensions for several samples of pre- and in-service teachers.

Second, this study did not provide specific information on which of the T-dimensions in the TPACK framework matter the most for integrating digital technologies and promoting students' digital literacy by creating meaningful and cognitively activating learning environments. In this respect, the role of the specific subjects still needs to be disentangled. Together with Baumert et al. (2010), we argue that the question what matters for student learning and progress is still to be determined.

Conclusion

The results of this study indicated that the measure of the technology-dimensions within the TPACK framework is able to capture pre-service teachers' general TPACK self-beliefs and their specific self-beliefs in technological knowledge. But given the high correlations among the pedagogical dimensions (i.e., TCK, TPCK, and TPK), the measure could not disentangle four separate factors. This result points to the question whether or not pre-service teachers' self-efficacy beliefs in these dimensions are actually distinguishable. Because this finding could be replicated across gender and educational track groups, it seems to persist. Nevertheless, TK represents a unique dimension among the T-dimensions. Looking at teacher education, we argue that it may be worthwhile fostering explicitly pre-service teachers' self-beliefs in their TK, and their self-beliefs in the more pedagogically oriented T-dimensions. In addition, the substantial differences in self-beliefs across gender call for action to help both female and male pre-service teachers to strengthen their self-beliefs and, at the same time, to help them develop a reasonable and accurate estimation of their abilities in the context of teaching with digital technologies. Finally, we point out that this study specifically adds to the existing body of research on the validity of TPACK measures by providing empirical evidence on their factor structure and generalizability across subgroups.

References

Abbitt, J. T. (2011). An Investigation of the Relationship between Self-Efficacy Beliefs about Technology Integration and Technological Pedagogical Content Knowledge (TPACK) among Preservice Teachers. *Journal of Digital Learning in Teacher Education, 27*(4), 134-143. doi:10.1080/21532974.2011.10784670

AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association (AERA).

Antonietti, A., & Giorgetti, M. (2006). Teachers' beliefs about learning from multimedia. *Computers in Human Behavior, 22*(2), 267-282. doi:10.1016/j.chb.2004.06.002

Archambault, L. M., & Barnett, J. H. (2010). Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education, 55*(4), 1656-1662. doi:10.1016/j.compedu.2010.07.009

Asparouhov, T., & Muthén, B. (2010). *Weighted least squares estimation with missing data*. Los Angeles, CA: Muthén & Muthén.

Ay, Y., Karadağ, E., & Acat, M. B. (2015). The Technological Pedagogical Content Knowledge-practical (TPACK-Practical) model: Examination of its validity in the Turkish culture via structural equation modeling. *Computers & Education, 88*, 97-108. doi:10.1016/j.compedu.2015.04.017

Baek, Y., Jung, J., & Kim, B. (2008). What makes teachers use technology in the classroom? Exploring the factors affecting facilitation of technology with a Korean sample. *Computers & Education, 50*(1), 224-234. doi:10.1016/j.compedu.2006.05.002

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman.

Banoglu, K., Vanderlinde, R., & Yildiz, R. (2015). Professional self-efficacy scale for information and computer technology teachers: validity and reliability study. *Anthropologist, 20*(1-2), 22-32.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal, 47*(1), 133-180. doi:10.3102/0002831209345157

Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 186-203. doi:10.1207/s15328007sem1302_2

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Dordrecht: Springer.

Blömeke, S., Suhl, U., Kaiser, G., & Döhrmann, M. (2012). Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education, 28*(1), 44-55. doi:10.1016/j.tate.2011.08.006

Bong, M., & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review, 15*(1), 1-40. doi:10.1023/A:1021302408382

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2 ed.). New York, NY: Guilford Press.

Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2010). Facilitating Preservice Teachers' Development of Technological, Pedagogical, and Content Knowledge (TPACK). *Educational Technology & Society, 13*(4), 63-73.

Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2013). A Review of Technological Pedagogical Content Knowledge. *Educational Technology & Society, 16*(2), 31-51.

Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2016). A review of the quantitative measures of Technological Pedagogical Content Knowledge (TPACK). In M. C. Herring, M. J. Koehler, & P. Mishra (Eds.), *Handbook of Technological Pedagogical Content Knowledge (TPACK) for Educators* (2 ed., pp. 87-106). New York, London: Taylor & Francis.

Chai, C. S., Koh, J. H. L., Tsai, C. C., & Tan, L. L. W. (2011). Modeling primary school pre-service teachers' Technological Pedagogical Content Knowledge (TPACK) for meaningful learning with information and communication technology (ICT). *Computers & Education, 57*(1), 1184-1193. doi:10.1016/j.compedu.2011.01.007

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. doi:10.1080/10705510701301834

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research, 41*(2), 189-225. doi:10.1207/s15327906mbr4102_5

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review, 15*, 1-16. doi:10.1016/j.edurev.2015.02.001

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. doi:10.1207/S15328007SEM0902_5

Chou, C.-P., & Huh, J. (2012). Model modification in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 232-246). New York, NY: The Guildford Press.

Compton, D. M., Burkett, W. H., & Burkett, G. G. (2003). No sex differences in perceived competence of computer use among male and female college students in 2002. *Psychological Reports, 92*(2), 503-511. doi:10.2466/pr0.2003.92.2.503

Durndell, A., & Haag, Z. (2002). Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior, 18*(5), 521-535. doi:10.1016/S0747-5632(02)00006-7

Eid, M., Geiser, C., Koch, T., & Heene, M. (2016). Anomalous Results in G-Factor Models: Explanations and Alternatives. *Psychological Methods*, No Pagination Specified. doi:10.1037/met0000083

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York, NY: Guilford Press.

Enders, C. K., & Mansolf, M. (2016). Assessing the Fit of Structural Equation Models With Multiply Imputed Data. *Psychological Methods*, No Pagination Specified. doi:10.1037/met0000102

Fisser, P., Voogt, J., Tondeur, J., & van Braak, J. (2015). Measuring and Assessing TPACK (technological, pedagogical, and content knowledge). In J. M. Spector (Ed.), *Encyclopaedia of Educational Technology* (pp. 490-493). Thousand Oaks, CA: SAGE Publications.

Fisser, P., Voogt, J., Van Braak, J., & Tondeur, J. (2013). *Unraveling the TPACK model: Finding TPACK-core*. Paper presented at the Society for Information Technology & Teacher Education International Conference, New Orleans, Louisiana, USA.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for Life in a Digital Age - The IEA International Computer and Information Literacy Study International Report*. Heidelberg: Springer International Publishing.

Gignac, G. E. (2009). Psychometrics and the Measurement of Emotional Intelligence. In A. J. D. Parker, H. D. Saklofske, & C. Stough (Eds.), *Assessing Emotional Intelligence: Theory, Research, and Applications* (pp. 9-40). Boston, MA: Springer US.

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education, 57*(3), 1953-1960. doi:10.1016/j.compedu.2011.04.010

Graham, C. R., Borup, J., & Smith, N. B. (2012). Using TPACK as a framework to understand teacher candidates' technology integration decisions. *Journal of Computer Assisted Learning, 28*(6), 530-546. doi:10.1111/j.1365-2729.2011.00472.x

Grossman, P. L., & Richert, A. E. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education, 4*(1), 53-62. doi:10.1016/0742-051X(88)90024-8

Guenole, N., & Brown, A. (2014). The Consequences of Ignoring Measurement Invariance for Path Coefficients in Structural Equation Models. *Frontiers in Psychology, 5*(980). doi:10.3389/fpsyg.2014.00980

He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response Styles and Personality Traits: A Multilevel Analysis. *Journal of Cross-Cultural Psychology, 45*(7), 1028-1045. doi:10.1177/0022022114534773

Hew, K. F., & Brush, T. (2006). Integrating technology into K-12 teaching and learning: current knowledge gaps and recommendations for future research. *Educational Technology Research and Development, 55*(3), 223-252. doi:10.1007/s11423-006-9022-5

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774-786. doi:10.1037/a0032198

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology, 41*, 13-24. doi:10.1016/j.cedpsych.2014.11.002

Jen, T.-H., Yeh, Y.-F., Hsu, Y.-S., Wu, H.-K., & Chen, K.-M. (2016). Science teachers' TPACK-Practical: Standard-setting using an evidence-based approach. *Computers & Education, 95*, 45-62. doi:10.1016/j.compedu.2015.12.009

Kaya, S., & Dağ, F. (2013). Turkish adaptation of Technological Pedagogical Content Knowledge Survey for Elementary Teachers. *Educational Sciences: Theory & Practice, 13*(1), 302-306.

Khojasteh, J., & Lo, W.-J. (2015). Investigating the Sensitivity of Goodness-of-Fit Indices to Detect Measurement Invariance in a Bifactor Model. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 531-541. doi:10.1080/10705511.2014.937791

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741-756. doi:10.1037/a0019237

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59-76. doi:10.1016/j.edurev.2014.06.001

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' Content Knowledge and Pedagogical Content Knowledge: The Role of Structural Differences in Teacher Education. *Journal of Teacher Education, 64*(1), 90-106. doi:10.1177/0022487112460398

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Koch, S. C., Müller, S. M., & Sieverding, M. (2008). Women and computers. Effects of stereotype threat on attribution of failure. *Computers & Education, 51*(4), 1795-1803. doi:10.1016/j.compedu.2008.05.007

Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The Technological Pedagogical Content Knowledge Framework. In M. J. Spector, D. M. Merrill, J. Elen, & J. M. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 101-111). New York, NY: Springer.

Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches* (pp. 16-31). Hersey, PA: IGI Global.

Koh, J. H. L., Chai, C. S., & Tay, L. Y. (2014). TPACK-in-Action: Unpacking the contextual influences of teachers' construction of technological pedagogical content knowledge (TPACK). *Computers & Education, 78*, 20-29. doi:10.1016/j.compedu.2014.04.022

Koh, J. H. L., Chai, C. S., & Tsai, C.-C. (2010). Examining the technological pedagogical content knowledge of Singapore pre-service teachers with a large-scale survey. *Journal of Computer Assisted Learning, 26*(6), 563-573.

Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2013). Examining practicing teachers' perceptions of technological pedagogical content knowledge (TPACK) pathways: A structural equation modeling approach. *Instructional Science, 41*(4), 793-809. doi:10.1007/s11251-012-9249-y

Kopcha, T. J., Ottenbreit-Leftwich, A., Jung, J., & Baser, D. (2014). Examining the TPACK framework through the convergent and discriminant validity of two measures. *Computers & Education, 78*, 87-96. doi:10.1016/j.compedu.2014.05.003

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: validation of the COACTIV constructs. *ZDM, 40*(5), 873-892. doi:10.1007/s11858-008-0141-9

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., . . . Elsner, J. (2013). Mathematics Teachers' Domain-Specific Professional Knowledge: Conceptualization and Test Construction in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (pp. 147-174). Boston, MA: Springer US.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716-725. doi:10.1037/0022-0663.100.3.716

Lei, P.-W., & Wu, Q. (2012). Estimation in Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164-180). New York, NY: Guilford Press.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary Psychometrics* (Vol. 275-340). Mahwah, NJ: Lawrence Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320-341. doi:10.1207/s15328007sem1103_2

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110. doi:10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 439-476. doi:10.1080/10705510903008220

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592. doi:10.1037/0021-9010.93.3.568

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. doi:10.1037/0003-066X.50.9.741

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054. doi:10.1111/j.1467-9620.2006.00684.x

Muthén, B., & Muthén, L. (1998-2015). MPlus (Version 7.3). Los Angeles, CA: Muthén & Muthén.

OECD. (2014). *TALIS 2013 Results*. Paris: OECD Publishing.

Olofson, M. W., Swallow, M. J. C., & Neumann, M. D. (2016). TPACKing: A constructivist framing of TPACK to analyze teachers' construction of knowledge. *Computers & Education, 95*, 188-201. doi:10.1016/j.compedu.2015.12.010

Pamuk, S., & Peker, D. (2009). Turkish pre-service science and mathematics teachers' computer related self-efficacies, attitudes, and the relationship between these variables. *Computers & Education, 53*(2), 454-461. doi:10.1016/j.compedu.2009.03.004

Paulick, I., Großschedl, J., Harms, U., & Möller, J. (2016). Preservice Teachers' Professional Knowledge and Its Relation to Academic Self-Concept. *Journal of Teacher Education*. doi:10.1177/0022487116639263

Plumm, K. M. (2008). Technology in the classroom: Burning the bridges to the gaps in gender-biased education? *Computers & Education, 50*(3), 1052-1068. doi:10.1016/j.compedu.2006.10.005

Price, L. (2017). *Psychometric Methods*. New York, NY: The Guilford Press.

Revelle, W., & Zinbarg, R. E. (2008). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145-154. doi:10.1007/s11336-008-9102-z

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373. doi:10.1037/a0029315

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices. *Psychological Methods*. doi:10.1037/met0000045

Rutkowski, L., & Svetina, D. (2014). Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys. *Educational and Psychological Measurement, 74*(1), 31-57. doi:10.1177/0013164413498257

Sahin, I. (2011). Development of survey of Technological Pedagogical and Content Knowledge (TPACK). *The Turkish Online Journal of Educational Technology, 10*(1), 97-105.

Sang, G., Tondeur, J., Chai, C. S., & Dong, Y. (2016). Validation and profile of Chinese pre-service teachers' technological pedagogical content knowledge scale. *Asia-Pacific Journal of Teacher Education, 44*(1), 49-65. doi:10.1080/1359866X.2014.960800

Sang, G., Valcke, M., Braak, J. v., & Tondeur, J. (2010). Student teachers' thinking processes and ICT integration: Predictors of prospective teaching behaviors with educational technology. *Computers & Education, 54*(1), 103-112. doi:10.1016/j.compedu.2009.07.010

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika, 75*(2), 243-248. doi:10.1007/s11336-009-9135-y

Scherer, R., Jansen, M., Nilsen, T., Areepattamannil, S., & Marsh, H. W. (2016). The Quest for Comparability: Studying the Invariance of the Teachers' Sense of Self-Efficacy (TSES) Measure across Countries. *PLoS One, 11*(3). doi:10.1371/journal.pone.0150829

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7*(110). doi:10.3389/fpsyg.2016.00110

Scherer, R., & Siddiq, F. (2015). Revisiting teachers' computer self-efficacy: A differentiated view on gender differences. *Computers in Human Behavior, 53*, 48-57. doi:10.1016/j.chb.2015.06.038

Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education, 88*(0), 202-214. doi:10.1016/j.compedu.2015.05.005

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK). *Journal of Research on Technology in Education, 42*(2), 123-149. doi:10.1080/15391523.2009.10782544

Schmidt, W. H., Cogan, L., & Houang, R. (2011). The Role of Opportunity to Learn in Teacher Preparation: An International Context. *Journal of Teacher Education, 62*(2), 138-153.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445. doi:10.1037/1082-989X.7.4.422

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15*(2), 4-14. doi:10.3102/0013189x015002004

Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review, 57*(1), 1-23. doi:10.17763/haer.57.1.j463w79r56455411

Sieverding, M., & Koch, S. C. (2009). (Self-)Evaluation of computer competence: How gender matters. *Computers & Education, 52*(3), 696-701. doi:10.1016/j.compedu.2008.11.016

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611-625. doi:10.1037/0022-0663.99.3.611

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2 ed.). London: Sage.

Tatto, M. T., Peck, R., Schwille, J., Bankov, K., Senk, S. L., Rodriguez, M. C., . . . Rowley, G. (2012). *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education, 57*(4), 2432-2440. doi:10.1016/j.compedu.2011.06.008

Teo, T., Lee, C. B., Chai, C. S., & Wong, S. L. (2009). Assessing the intention to use technology among pre-service teachers in Singapore and Malaysia: A multigroup invariance analysis of the Technology Acceptance Model (TAM). *Computers & Education, 53*(3), 1000-1009. doi:10.1016/j.compedu.2009.05.017

Tondeur, J., van Braak, J., Siddiq, F., & Scherer, R. (2016). Time for a new approach to prepare future teachers for educational technology use: Its meaning and measurement. *Computers & Education, 94*, 134-150. doi:10.1016/j.compedu.2015.11.009

Tondeur, J., Van de Velde, S., Vermeersch, H., & Van Houtte, M. (2016). Gender Differences in the ICT Profile of University Students: A Quantitative Analysis. *Journal of Diversity and Gender Studies*.

Tondeur, J., van Keer, H., van Braak, J., & Valcke, M. (2008). ICT integration in the classroom: Challenging the potential of a school policy. *Computers & Education, 51*(1), 212-223. doi:10.1016/j.compedu.2007.05.003

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*(6), 944-956. doi:10.1016/j.tate.2006.05.003

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research, 68*(2), 202-248. doi:10.3102/00346543068002202

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education, 27*(4), 751-761. doi:10.1016/j.tate.2010.12.005

Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research, 78*(4), 751-796. doi:10.3102/0034654308321456

Vekiri, I., & Chronaki, A. (2008). Gender issues in technology use: Perceived social support, computer self-efficacy and value beliefs, and computer use beyond school. *Computers & Education, 51*(3), 1392-1404. doi:10.1016/j.compedu.2008.01.003

Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge – a review of the literature. *Journal of Computer Assisted Learning, 29*(2), 109-121. doi:10.1111/j.1365-2729.2012.00487.x

Voogt, J., Fisser, P., van Braak, J., & Tondeur, J. (2013). *Unraveling the TPACK model: finding TPACK-Core*. Paper presented at the Annual Conference of the Society for Information Technology and Teacher Education (SITE), New Orleans.

Zee, M., de Jong, P. F., & Koomen, H. M. Y. (2016). Teachers' Self-Efficacy in Relation to Individual Students With a Variety of Social–Emotional Behaviors: A Multilevel Investigation. *Journal of Educational Psychology*. doi:10.1037/edu0000106

Zumbo, B. D. (2006). Validity: Foundational Issues and Statistical Methodology. In C. R. Rao & S. Sinharay (Eds.), *Psychometrics* (Vol. 26, pp. 45-79). Amsterdam: Elsevier.

Tables

Table 1

Descriptive statistics and reliabilities of the T-dimensions measure (Dutch version and English translation)

Item	*Wording*	*M*	*SD*	α	ω
Technological content knowledge (TCK)					
TCK1	Ik ben op de hoogte van ICT-toepassingen die ik kan gebruiken om leerlingen inzicht te geven in het vakgebied. *I am aware of ICT applications that I can use to give students insight into the subject I teach.*	2.47	0.86	.89	.85
TCK2	Ik ben op de hoogte van ICT-toepassingen om het vakgebied te ondersteunen. *I am aware of ICT applications to support the subject I teach.*	2.49	0.86		
TCK3	Ik kan ICT-toepassingen kiezen die lessen in een vakgebied ondersteunen. *I can choose ICT applications that support lessons a subject domain.*	2.74	0.75		
TCK4	Ik weet hoe ik ICT-toepassingen kan gebruiken om concepten uit een vakgebied op een andere manier te presenteren aan mijn leerlingen. *I know how to use ICT applications to present concepts from a discipline in a different way to my students.*	2.67	0.78		
Technological pedagogical knowledge (TPK)					
TPK1	Ik ben in staat ICT-toepassingen te kiezen die het leerproces van de leerlingen versterken. *I can choose technologies that enhance students' learning for a lesson.*	2.67	0.76	.87	.84
TPK2	Ik ben in staat ICT-toepassingen te kiezen die didactische werkvormen voor een les versterken. *I can choose technologies that enhance the teaching approaches for a lesson.*	2.77	0.72		
TPK3	Ik kan mijn ICT-gebruik afstemmen op verschillende leeractiviteiten. *I can adapt the use of the technologies that I am learning about to different teaching activities.*	2.67	0.76		
TPK4	Ik denk kritisch na over de manier waarop ik ICT-toepassingen in mijn eigen klas kan gebruiken. *I am thinking critically about how to use technology in my classroom.*	2.77	0.73		
TPK5	Door mijn opleiding denk ik kritisch na over de manier waarop ICT mijn didactische aanpak in	2.51	0.93		

	de klas kan beïnvloeden. *My teacher education program has caused me to think more deeply about how technology could influence the teaching approaches I use in my classroom.*				
Technological pedagogical content knowledge (TPCK)					
TPCK1	Ik kan lessen geven waarbij ICT, vakinhoud en didactiek op een goede manier zijn geïntegreerd. *I can teach lessons that appropriately combine technologies, literacy, and teaching approaches.*	2.68	0.78	.92	.87
TPCK2	Ik kan strategieën die ik heb geleerd (bv in mijn opleiding) gebruiken in mijn lessen om vakinhoud, ICT en didactiek te combineren. *I can use strategies that I have learned (in my teacher education program) to combine ICT, content, and pedagogy.*	2.58	0.81		
TPCK3	Ik kan ICT-toepassingen kiezen die versterken wat en hoe ik onderwijs geef. *I can choose ICT applications that enhance what and how I teach.*	2.70	0.71		
TPCK4	Ik kan ICT-toepassingen kiezen voor een vakgebied die versterken wat en hoe ik onderwijs geef. *I can choose ICT applications for a subject domain that enhance what and how I teach.*	2.66	0.76		
TPCK5	Ik kan lessen geven over een vakgebied waarbij ICT, vakinhoud en didactiek op een juiste manier zijn geïntegreerd. *I can give lessons about a subject area that appropriately integrate ICT, content, and teaching approaches.*	2.61	0.80		
Technological knowledge (TK)					
TK1	Ik kan mijn eigen ICT-problemen oplossen. *I know how to solve my own technical problems.*	2.47	0.96	.92	.85
TK2	Ik leer gemakkelijk nieuwe dingen over ICT. *I can learn technology easily.*	2.66	0.91		
TK3	Ik blijf op de hoogte van belangrijke ICT-ontwikkelingen. *I keep up with important new technologies.*	2.18	0.99		
TK4	Ik probeer regelmatig dingen uit met ICT. *I frequently play around with the technology.*	2.58	0.92		
TK5	Ik ken veel verschillende ICT-toepassingen. *I know about a lot of different technologies.*	2.47	0.95		
TK6	Ik beschik over de technische vaardigheden die ik nodig heb om ICT te gebruiken. *I have the technical skills I need to use technology.*	2.72	0.84		
TK7	Ik heb voldoende mogelijkheden om	2.51	0.90		

	verschillende ICT-toepassingen te gebruiken. *I have had sufficient opportunities to work with different technologies.*	

Note. α = Cronbach's α, ω = McDonald's ω.

Table 2

Fit statistics of T-dimensions measurement models for the total sample (N = 665)

Model	*SB-* χ^2 *(df)*	*CFI*	*TLI*	*RMSEA (90% CI)*	*SRMR*	*AIC*	*BIC*	*aBIC*	*Reference model*	*SB-* $\Delta\chi^2$ *(* Δdf *)*
Single-factor model	670.8 (187)*	.919	.909	.062 (.057, .068)	.049	26420	26712	26507	–	–
Four-factor model	372.2 (181)*	.968	.963	.040 (.034, .046)	.038	25980	26300	26074	Single-factor model	206.5 (6)*
Two-factor model	388.6 (186)*	.966	.962	.040 (.035, .046)	.039	25991	26288	26078	Single factor model	52.3 (1)*
Bifactor model[#]	270.2 (168)*	.983	.979	.030 (.023, .037)	.027	25861	26239	25972	Four-factor model	105.1 (13)*
Nested factor model	311.0 (180)*	.978	.974	.033 (.027, .039)	.028	25893	26217	25989	Four-factor model	43.7 (1)*

Note. SB- χ^2 refers to the Satorra-Bentler corrected χ^2 value (Satorra & Bentler, 2010). AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, aBIC = sample-size adjusted BIC. # In order to achieve convergence, residual correlations were not specified in this model.

* $p < .01$

Table 3

Factor correlations among the four T-dimensions for the total sample

	TCK	TPK	TPCK	TK
TCK	1.00			
TPK	.98 (.02)	1.00		
TPCK	.98 (.01)	.99 (.02)	1.00	
TK	.86 (.04)	.81 (.04)	.81 (.04)	1.00

Note. *N* = 665. Standard errors are shown in parentheses. All correlations are statistically significant at the 1% level.

Table 4

Factor loadings and specificities in the nested factor model (total sample)

Items	*Factor loadings*		*Specificity*
	General factor λ_{GEN}	*Specific TK factor* λ_{TK}	
TCK1	.758 (.020)	–	–
TCK2	.752 (.031)	–	–
TCK3	.761 (.019)	–	–
TCK4	.750 (.022)	–	–
TPK1	.781 (.028)	–	–
TPK2	.767 (.028)	–	–
TPK3	.758 (.028)	–	–
TPK4	.625 (.029)	–	–
TPK5	.546 (.049)	–	–
TPCK1	.756 (.029)	–	–
TPCK2	.704 (.046)	–	–
TPCK3	.779 (.023)	–	–
TPCK4	.773 (.026)	–	–
TPCK5	.793 (.020)	–	–
TK1	.528 (.051)	.560 (.042)	.529
TK2	.567 (.045)	.525 (.038)	.462
TK3	.631 (.026)	.365 (.043)	.251
TK4	.573 (.045)	.414 (.047)	.343
TK5	.654 (.025)	.424 (.034)	.296
TK6	.666 (.048)	.435 (.062)	.299
TK7	.695 (.036)	.132 (.036)	.035

Note. All factor loadings are standardized and statistically significant at the 1% level.

Table 5

Measurement invariance of the nested factor model describing the structure of the T-dimensions across gender and educational tracks

Model	SB-χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR	*SB*-$\Delta\chi^2$ (Δ*df*)	ΔCFI	ΔRMSEA	ΔSRMR
Grouping variable: Gender									
Configural	591.7 (360)*	.965	.959	.044 (.038, .050)	.036	–	–	–	–
Metric	621.1 (386)*	.964	.961	.043 (.037, .049)	.046	29.6 (26), $p = .28$	–.001	–.001	+.010
Scalar	666.7 (405)*	.960	.959	.044 (.038, .050)	.050	75.0 (45), $p < .01$	–.005	.000	+.014
Grouping variable: Educational track									
Configural	610.2 (360)*	.966	.960	.046 (.039, .052)	.036	–	–	–	–
Metric	632.6 (386)*	.966	.963	.044 (.038, .050)	.043	21.0 (26), $p = .74$	.000	–.002	+.007
Scalar	651.6 (405)*	.966	.965	.043 (.037, .049)	.044	41.5 (45), $p = .62$	.000	–.003	+.008

Note. SB-χ^2 refers to the Satorra-Bentler corrected χ^2 value (Satorra & Bentler, 2010).

Figures

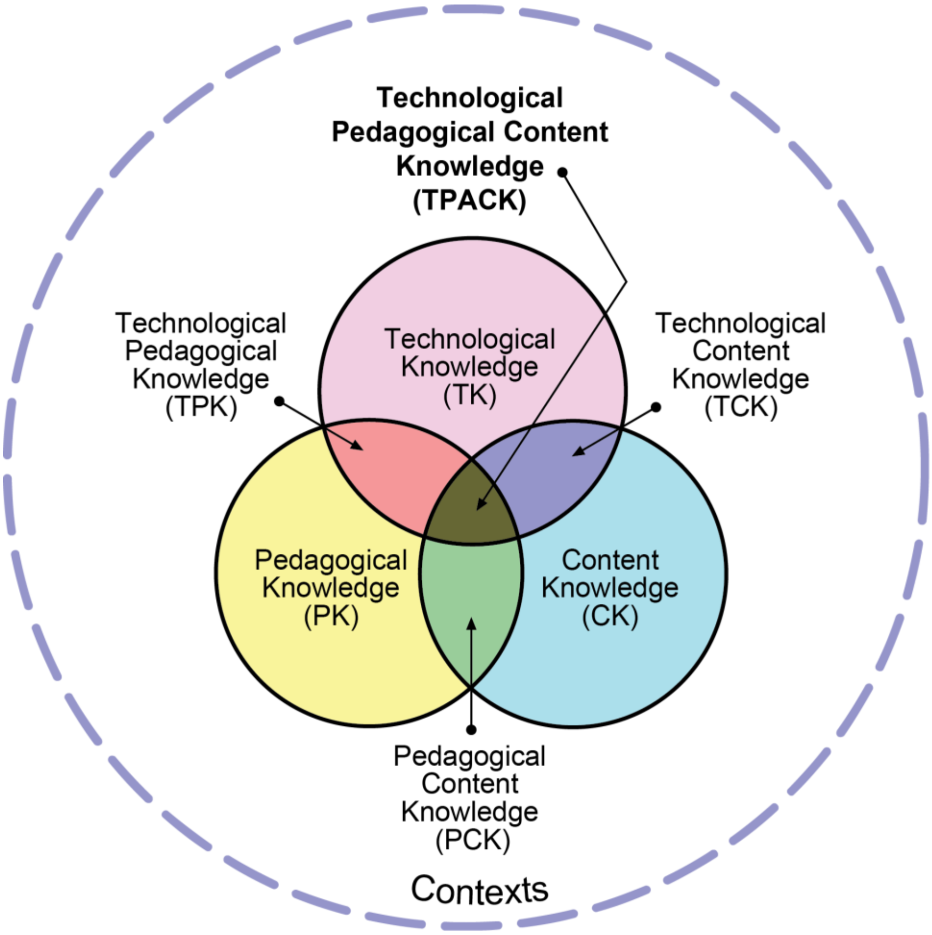

Figure 1. The TPACK model.

Note. Reproduced by permission of the publisher, © 2012 by tpack.org.

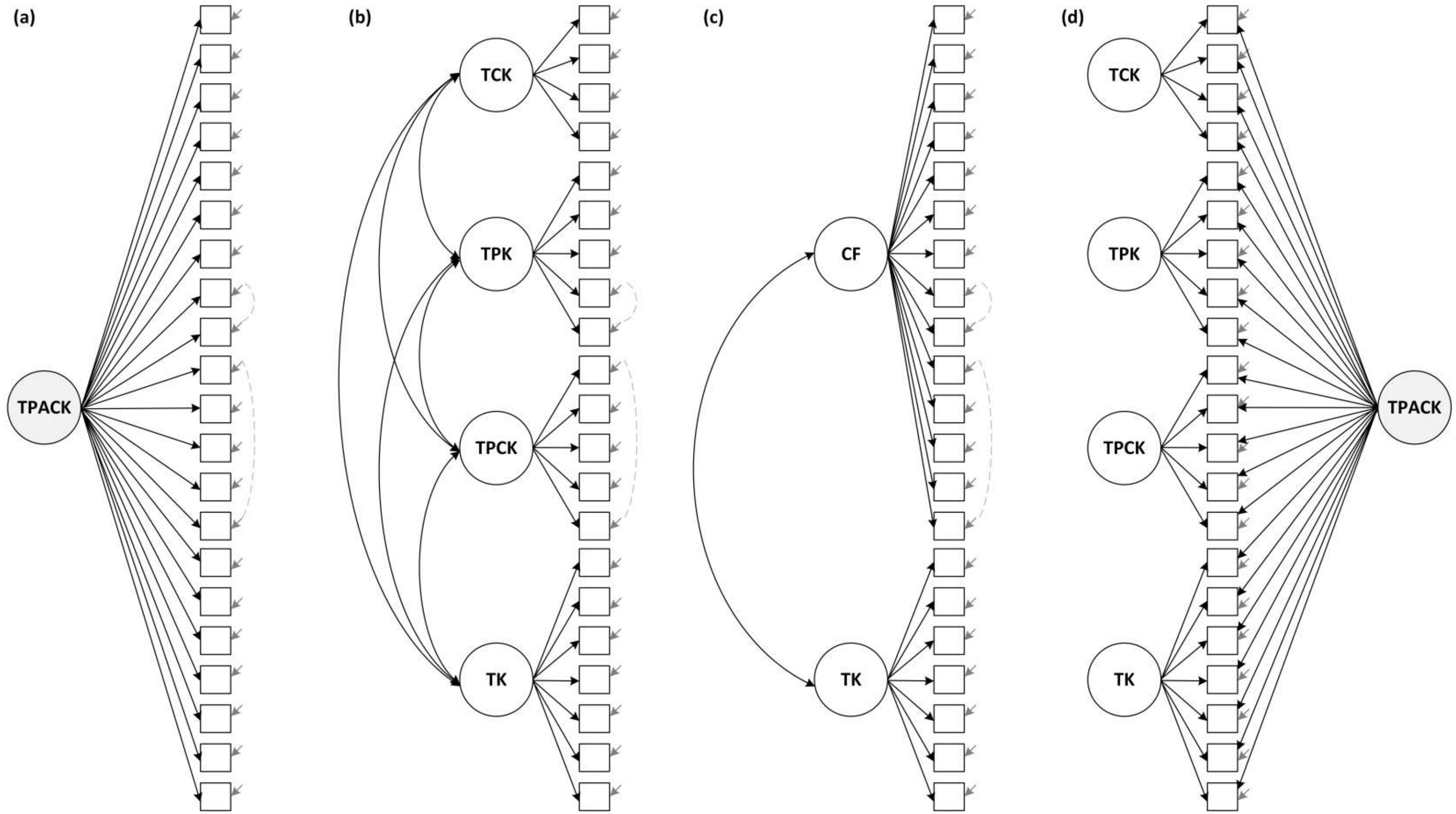

Figure 2. Theoretically implied measurement models of the T-dimensions: (a) Model assuming a single TPACK factor; (b) Four-correlated-factors model; (c) Two-correlated-factors model; (d) Bifactor model comprising a general TPACK factor and four uncorrelated, specific factors

Note. TCK = Technological content knowledge, TPK = Technological pedagogical knowledge, TPCK = Technological pedagogical content knowledge, TK = Technological knowledge, CF = Combined factor (TCK, TPK, and TPCK).

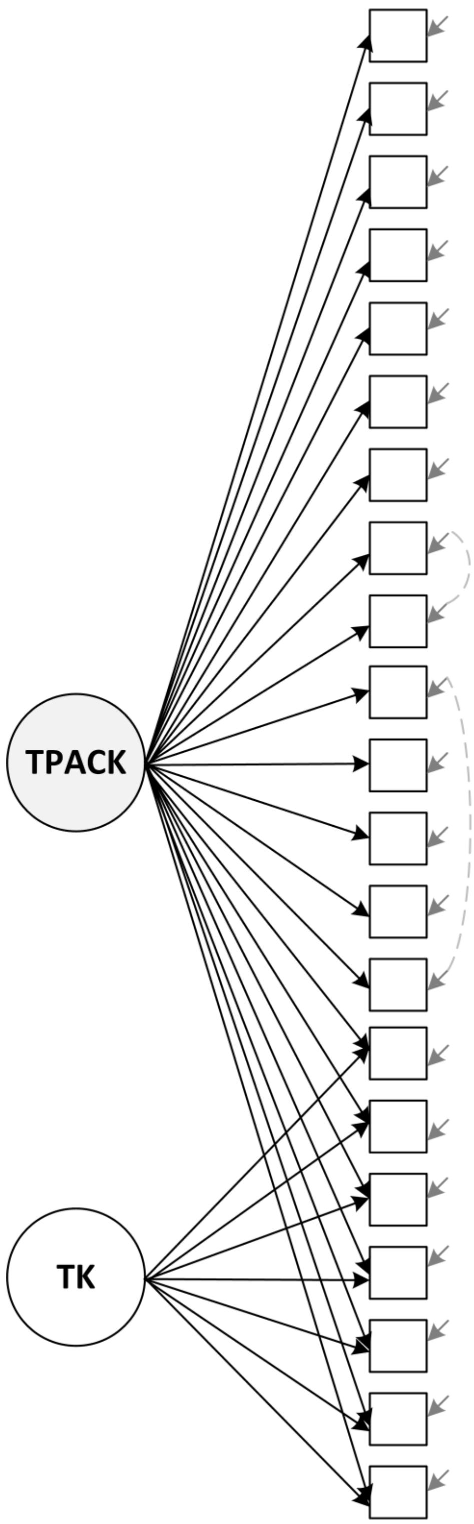

Figure 3. Empirical measurement model of the T-dimensions

Note. Residual correlations are indicated by dashed lines. TK = Technological knowledge.

Supplementary Material

On the Quest for Validity: Testing the Factor Structure and Measurement Invariance of the Technology-Dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) Model

A) Exploratory structural equation modeling

Table A1

Standardized factor loadings of the two-factor ESEM solution

Items	*Factor 1*	*Factor 2*
TCK1	**.465 (.031)***	.390 (.035)*
TCK2	**.455 (.054)***	.394 (.052)*
TCK3	**.701 (.042)***	.115 (.042)*
TCK4	**.597 (.049)***	.222 (.047)*
TPK1	**.628 (.047)***	.225 (.042)*
TPK2	**.695 (.039)***	.131 (.038)*
TPK3	**.633 (.044)***	.192 (.038)*
TPK4	**.495 (.044)***	.191 (.045)*
TPK5	**.423 (.075)***	.172 (.058)*
TPCK1	**.564 (.039)***	.271 (.045)*
TPCK2	**.546 (.070)***	.223 (.063)*
TPCK3	**.691 (.027)***	.149 (.030)*
TPCK4	**.667 (.038)***	.169 (.043)*
TPCK5	**.672 (.042)***	.187 (.043)*
TK1	-.075 (.037)	**.795 (.034)***
TK2	-.013 (.039)	**.767 (.028)***
TK3	.123 (.041)*	**.666 (.043)***
TK4	.071 (.046)	**.663 (.041)***
TK5	.109 (.032)*	**.719 (.035)***
TK6	.144 (.055)*	**.695 (.043)***
TK7	**.416 (.046)***	.379 (.045)*

Note. Standard errors are shown in parentheses; the highest factor loadings are shown in bold.
* $p < .01$

Table A2

Standardized factor loadings of the three-factor ESEM solution

Items	*Factor 1*	*Factor 2*	*Factor 3*
TCK1	.235 (.032)*	**.469 (.052)***	.199 (.051)*
TCK2	.139 (.070)	**.620 (.101)***	.132 (.078)*
TCK3	**.668 (.053)***	.067 (.048)	.158 (.043)*
TCK4	**.466 (.056)***	.250 (.072)*	.163 (.063)*
TPK1	**.506 (.067)***	.234 (.058)*	.176 (.036)*
TPK2	**.643 (.023)***	.095 (.051)	.164 (.056)*
TPK3	**.519 (.038)***	.211 (.057)*	.159 (.044)*
TPK4	**.381 (.061)***	.217 (.072)*	.137 (.047)*
TPK5	.161 (.116)	**.481 (.113)***	-.024 (.038)
TPCK1	**.441 (.066)***	.246 (.092)*	.210 (.047)*
TPCK2	.262 (.098)*	**.524 (.085)***	.021 (.038)
TPCK3	**.525 (.043)***	.285 (.066)*	.091 (.040)
TPCK4	**.535 (.026)***	.235 (.035)*	.130 (.053)
TPCK5	**.485 (.054)***	.329 (.052)*	.103 (.031)*
TK1	.003 (.044)	.012 (.055)	**.772 (.045)***
TK2	.050 (.049)	.040 (.063)	**.729 (.043)***
TK3	-.202 (.045)*	**.667 (.096)***	.357 (.091)*
TK4	.033 (.027)	.213 (.053)*	**.537 (.052)***
TK5	-.017 (.034)	.371 (.060)*	**.522 (.048)***
TK6	.185 (.046)*	.057 (.046)	**.672 (.054)***
TK7	.218 (.056)*	**.410 (.071)***	.213 (.044)*

Note. Standard errors are shown in parentheses; the highest factor loadings are shown in bold.
* $p < .01$

Table A3

Standardized factor loadings of the four-factor ESEM solution

Items	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*
TCK1	.205 (.063)*	.185 (.219)	**.438 (.274)**	.109 (.070)
TCK2	.151 (.132)	.143 (.293)	**.617 (.411)**	.006 (.074)
TCK3	**.566 (.090)***	.197 (.176)	.045 (.223)	.131 (.114)
TCK4	.298 (.183)	**.389 (.286)**	.074 (.242)	.163 (.140)
TPK1	**.469 (.144)***	.131 (.128)	.254 (.118)	.104 (.054)
TPK2	**.564 (.069)***	.164 (.092)	.095 (.058)	.123 (.033)*
TPK3	**.482 (.085)***	.124 (.043)*	.232 (.084)*	.092 (.047)
TPK4	.237 (.100)*	**.328 (.150)**	.076 (.121)	.130 (.052)
TPK5	-.089 (.167)	**.605 (.161)***	.129 (.100)	.002 (.040)
TPCK1	.296 (.104)*	**.345 (.135)***	.103 (.138)	.198 (.092)
TPCK2	-.008 (.096)	**.652 (.307)**	.170 (.252)	.031 (.042)
TPCK3	**.467 (.174)***	.185 (.156)	.259 (.062)*	.031 (.049)
TPCK4	**.402 (.033)***	.308 (.062)*	.124 (.079)	.108 (.065)
TPCK5	.321 (.063)*	**.406 (.052)***	.141 (.051)*	.095 (.053)
TK1	-.018 (.137)	.051 (.057)	.065 (.109)	**.730 (.051)***
TK2	.043 (.069)	.034 (.096)	.118 (.131)	**.665 (.060)***
TK3	-.171 (.220)	.148 (.256)	**.630 (.306)**	.239 (.233)
TK4	.006 (.052)	.115 (.150)	.225 (.191)	**.471 (.111)***
TK5	.035 (.160)	.006 (.228)	**.466 (.272)**	.401 (.211)
TK6	.157 (.128)	.073 (.088)	.120 (.046)*	**.609 (.056)***
TK7	.177 (.067)*	.194 (.097)	**.363 (.105)***	.140 (.056)

Note. Standard errors are shown in parentheses; the highest factor loadings are shown in bold.
* $p < .01$

B) Supplementary statistics

Table B1

Summary statistics of the TPACK items (N = 665)

Item	*M*	*SD*	*Mdn*	*Min*	*Max*	*Skewness*	*Kurtosis*	*SE*
TPCK1	2.68	0.78	3	0	4	-0.92	1.33	0.03
TPCK2	2.58	0.81	3	0	4	-0.92	0.93	0.03
TPCK3	2.70	0.71	3	0	4	-0.85	1.30	0.03
TPCK4	2.66	0.76	3	0	4	-0.93	1.26	0.03
TPCK5	2.61	0.80	3	0	4	-0.73	0.81	0.03
TCK1	2.47	0.86	3	0	4	-0.62	0.16	0.03
TCK2	2.49	0.86	3	0	4	-0.65	0.19	0.03
TCK3	2.74	0.75	3	0	4	-1.06	1.87	0.03
TCK4	2.67	0.78	3	0	4	-0.90	1.14	0.03
TPK1	2.67	0.76	3	0	4	-0.89	1.10	0.03
TPK2	2.77	0.72	3	0	4	-1.08	2.37	0.03
TPK3	2.67	0.76	3	0	4	-0.91	1.04	0.03
TPK4	2.77	0.73	3	0	4	-0.78	1.20	0.03
TPK5	2.51	0.93	3	0	4	-0.79	0.38	0.04
TK1	2.47	0.96	3	0	4	-0.46	-0.23	0.04
TK2	2.66	0.91	3	0	4	-0.73	0.38	0.04
TK3	2.18	0.99	3	0	4	-0.10	-0.69	0.04
TK4	2.58	0.92	3	0	4	-0.57	0.03	0.04
TK5	2.47	0.95	3	0	4	-0.50	-0.12	0.04
TK6	2.72	0.84	3	0	4	-0.80	0.86	0.03
TK7	2.51	0.90	3	0	4	-0.74	0.23	0.04