Validation of the Survey of Pre-service Teachers' Knowledge of Teaching and Technology: A Multi-Institutional Sample

Albert D. Ritzhaupt, Anne Corinne Huggins-Manley, Krista Ruggles, & Matthew Wilson
University of Florida

Abstract

The TPACK (technological pedagogical content knowledge) framework (Mishra & Koehler, 2006) has gained tremendous momentum from within the educational technology community. Specifically, much discourse has focused on how to measure this multidimensional construct to further define the contours of the framework and potentially make some meaningful predictions. Some have proposed observation scales while other have proposed self-report measures to gauge the phenomenon. The Survey of Pre-service Teachers' Knowledge of Teaching and Technology instrument is one popular tool designed to measure TPACK (Schmidt et al., 2009) specifically from preservice teachers in teacher education programs. This study extends the measurement framework by providing a confirmatory factor analysis of the theoretical model proposed by Schmidt et al. (2009) on a sample of 227 preservice teachers from four public institutions of higher education in the southeastern United States. The data did not fit the theoretical 10-factor model implied by Schmidt et al. (2009), thus, an exploratory factor analysis was conducted to determine the optimal structure of the measurement tool for these data. This resulted in a nine-factor model, and there were measurement issues for several of the constructs. Additionally, the article provides evidence of external validity by correlating the instrument scores with other known technology constructs.

Color versions of one or more of the figures in the article can be found online at http://www.tandfonline.com/ujdl.

As with the development and advancement of virtually any theory or conceptual framework, we turn to measurement as a way to systematically and intentionally study the contours of the potential construct and make predictions about the phenomenon. There is little disagreement that TPACK (technological pedagogical content knowledge) has had a major influence on the discourse in educational technology research. As of July 2014, the seminal article by Mishra and Koehler (2006), titled "Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge," has garnered more than 2,000 citations according to Google Scholar. Further, the term *TPACK* can be found in the title of more than a thousand publications to date (GoogleScholar, 2014). As we have struggled with the discourse on technology integration practices of teachers for decades, the TPACK framework provides a much needed framework to study the knowledge of teachers involved with technology integration through a meaningful lens.

With the development of the TPACK framework and the discourse that soon followed, it is no surprise that measurement tools have been developed in the research base to measure this multidimensional construct and make sense of teacher technology integration knowledge and practices. Perhaps the most influential tool to date is the Survey of Pre-service Teachers' Knowledge of Teaching and Technology (TKTT), developed by Schmidt, Baran, Thompson, Mishra, Koehler, and Shin (2009). While this measurement tool has been used widely in the research base (cited more than 250 times according to Google Scholar and used in several relevant publications on TPACK, e.g., Abbitt, 2011; Chai, Koh, & Tsai, 2010; Zellowski, Gleason, Cox, & Bismark, 2013), there remain questions about the stability and structure of the instrument (Chai, Koh, Tsai & Tan, 2011; Koh, Chai, Tsai, 2010). Thus, this article extends our understanding of the measurement of the multidimensional TPACK construct by empirically testing the aforementioned survey tool. First, we provide a review of the TPACK framework and examine some of the existing literature wherein studies attempt to operationalize and measure the TPACK framework.

TPACK Framework

To address the perceived need for both researchers and practitioners, Mishra and Koehler (2006) developed a framework to address the theory, empirical research, and practical application of educational technology. Shulman (1986) proposed the idea of pedagogical content knowledge (PCK) comprised of pedagogical knowledge (PK) and content knowledge (CK), wherein researchers and practitioners examine subject-matter content as "the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others" (p. 9). Founded upon Shulman's premise, Mishra and Koehler (2006) conceptualized the domain of technological knowledge (TK) to address the rise of technology use within the realm of teaching and learning. They propose that the technological knowledge is isolated from each pedagogical and content knowledge, but demands consideration in conjunction with the two (Mishra & Koehler, 2006). In the resulting framework, Mishra and Koehler (2006) name two new relatives of PCK, technological pedagogical knowledge (TPK) and technological content knowledge (TCK), as well as the cumulative intersection of all domains as TPCK, or technological pedagogical content knowledge, and later TPACK (Thompson & Mishra, 2007). Figure 1 visualizes the TPACK framework and the three intersecting areas.

TPACK in Use

Since its inception, researchers and practitioners have used the TPACK framework broadly to generate dialog and ideas toward technology integration in education. "Though not all teachers have embraced these new technologies for a range of reasons—including a fear of change and lack of time and support—the fact that these technologies are here to stay cannot be doubted" (Mishra & Koehler, 2006, p. 1023). In their conceptualization of TPACK, Mishra and Koehler (2006) outline two practical purposes of TPACK: the development of teacher understanding of integration, and framing research. These two avenues of application are often one and the same in literature. In a major cross section of TPACK studies, teacher education researchers use the framework to approach the question of technology integration for preservice teachers (Chai, Koh, & Tsai, 2010; Koh & Divaharan, 2011; Niess, 2005). Furthermore, researchers similarly apply TPACK in the professional development of in-service educators (Doering, Veletsianos, Scharber, & Miller, 2009; Graham et al., 2009). The ongoing examination and application of TPACK into practice and research have generated a dialog on the criticisms of the framework.

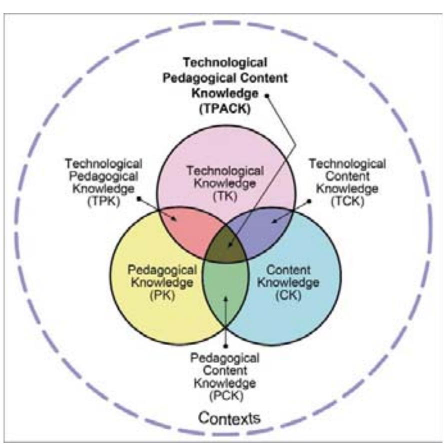

Figure 1. TPACK framework.

TPACK Criticisms and Extensions

Thompson and Schmidt (2010) propose that the formulation of the TPACK framework provides practitioners and researchers with a common vocabulary for discussing the idea of technology integration for advancing student learning. While TPACK helps fulfill that role, there are numerous critiques within current literature that highlight the shortcomings of the frame work, as well as some of its extensions.

Brantley-Dias and Ertmer (2013) criticize the TPACK framework for not clearly gauging what types of pedagogy or curricula provide a "best fit" for technology integration. Graham (2011) and Archambault and Barnett (2010) each call into question the theoretical foundations of TPACK by stating that Shulman's PCK operated with difficult-to-define domains that make the overall construct unclear:

> The TPACK framework builds on the PCK framework and increases the conceptual complexity by at least an order of magnitude. Because PCK is foundational to the TPACK framework, researchers must clearly understand PCK before they can productively understand and effectively measure TPACK constructs. (Graham, 2011, p. 1955)

Furthermore, Brantley-Dias and Ertmer (2013) state that TPACK possesses a critical flaw of being both too large (seven distinct knowledge types) and too small (compartmentalized) for practical use or measure. This dichotomy is central to the ongoing search for a viable TPACK evaluation tool.

Although several criticisms of the TPACK framework have been made, there have also been several useful extensions to this framework. TPACK has previously been used as a framework to support research on technology integration, including case studies of mathematics teachers involved in a learner-centered professional development project (Polly, 2009) and mathematics and science preservice teachers enrolled in methods courses (Niess, 2005), survey research to ascertain K–12 online teachers' perceptions of their TPACK knowledge (Archambault & Crippen, 2009) and to study how faculty and students develop TPACK in a learning technology by design seminar (Koehler & Mishra, 2005), interpretive research examining growth of TPCK knowledge exhibited by in-service teachers enrolled in an online graduate course (Niess, van Zee, Gillow-Wiles, 2010), and design-based research to support TPACK development in preservice teachers (Mishra & Koehler, 2006).

One extension worth noting is the connection of TPACK domains with various learning activity types for educational technologies by Harris, Mishra, and Koehler (2009). "Technologies' affordances create opportunities for both enhancing existing learning activity types and creating new ones" (Harris, Mishra, and Koehler, 2009, p. 406). Harris, Mishra, and Koehler (2009) provide examples of knowledge-building, convergent, and divergent activity types connected to TPACK to further define the constructs for TPACK users. This extension is particularly useful when thinking about possible activities for preservice teacher education programs that effectively integrate technology.

Measuring the TPACK Framework

The measurement and evaluation of the TPACK framework as a viable model for developing and assessing the effectiveness of technology integration have taken place since its inception (Mishra & Koehler, 2006). These studies have primarily focused on surveying methods of preservice and in-service teachers' self-assessment of technology use, understanding, or issues (Archambualt & Crippen, 2009; Keller, Bonk, & Hew, 2005) and teachers' change in the perceptions of their understanding (An, Wilder, & Lim, 2011; Koehler & Mishra, 2005; Koehler, Mishra, & Yahya, 2007). However, due to the framework being a relatively new model, empirical research focusing on existing measurement instruments' validity and reliability is limited (Young, Young, & Hamilton, 2013). The Survey of Pre-service Teachers' Knowledge of Teaching and Technology (TKTT) instrument was developed as part of a study that measured "pre-service teachers' self-assessment of the seven knowledge domains within the TPACK framework" (Schmidt et al., 2009, p. 128). The self-assessment instrument contained 75 items and prompted the participant to respond to each item using a five-level Likert scale (Schmidt et al., 2009). Participants included 124 elementary and early childhood preservice teachers enrolled in a semester-long technology integration course (Schmidt et al., 2009). The online survey was completed during their final class session (Schmidt et al., 2009). Due to the small sample size, Schmidt et al. (2009) recognized the limitation of their sample. After completing a factor analysis and using the Kaiser–Guttman rule, the researchers evaluated each item and eliminated those that posed problems in terms of validity or reliability (Schmidt et al., 2009). Notwithstanding the small sample size, the study found that the refined TKTT could serve as a potentially promising self-assessment tool for measuring preservice teachers' TPACK (Schmidt et al., 2009).

While multiple studies have utilized the Schmidt et al. TKTT survey as a means to measure preservice teacher TPACK, there are noted limitations. Abbitt's (2011) study of 45 preservice teachers used the TKTT to collect pre and post data over a 16-week technology integration course; however, the small and homogeneous student sample represented in this study cannot be generalized to a larger diverse population. Chai, Koh, and Tsai's (2010) study of preservice secondary teachers used the TKTT survey to analyze precourse and postcourse perceptions of TK, PK, and CK to TPACK. This voluntary study included 439 precourse sand 365 postcourse survey participants. While the study found somewhat large effect sizes in the TK, PK, and CK domains, the study did not include PCK, TCK, and TPK due to the course design (Chai, Koh & Tsai, 2010). Zellowski, Gleason, Cox, and Bismark's (2013) study used a modified TKTT survey focused on mathematics content areas. While the sample size of 315 preservice secondary teachers from 15 U.S. institutions represented a larger and more diverse student population, the results of the exploratory factor analysis (EFA) only presented specific factors for TK, CK, PK, and TPACK. The researchers selected to remove PCK, TCK, and TPK due to the absence of a clear pattern (Zellowski et al., 2013). Thus, more empirical research is needed on this measurement tool and its applications.

Purpose

Since Schmidt et al. (2009) developed the Survey of Pre-service Teachers' Knowledge of Teaching and Technology, some studies (Chai, Koh, & Tsai, 2010; Schmidt et al., 2009) have attempted to explore the internal structure of the instrument. However, these prior studies have primarily used an exploratory factor analysis to examine the optimal structure of the instrument as informed by observed data. The next logical step in an instrument development and validation process is to test whether the theoretical model from the Technological Pedagogical Content Knowledge (TPACK) framework fit the observed item response data. Thus, this article attempts to test the TPACK framework as a viable theoretical tool for measuring this phenomena. Additionally, we provide external validation of the tool by correlating meaningful factors structure with other technology measures theorized to have a relationship with TPACK. These two characteristics make our work distinct from prior research that has largely focused on the internal structure of the measurement system using an exploratory model, with the exception of Zelkowski, Gleason, Cox, and Bismarck (2013).

Method

Participants

Participants included $N = 227$ preservice elementary school teachers from four public universities in the southeastern United States enrolled in a technology integration course for teachers. Eighty-three percent of the participants were female, which is not unusual for elementary teacher education programs.

Table 1. Participant Distribution and Percentage by University Site and by Gender

University	*n*	Male (*n*)	Female (*n*)	Percent of total
University Site 1	74	8	66	35%
University Site 2	32	7	25	14%
University Site 3	48	12	36	20%
University Site 4	73	11	62	31%

Eighty-six percent of the participants were in the age range of 18 to 22 years old, and the remaining ones more than 23 years old. In terms of ethnicity, 76% of the participants self-identified as White, 7% as Black, 11% as Hispanic, 3% as Asian, and the remaining classified as Other. Table 1 provides the distribution of the participants by each university and by gender. All of the participants had recently completed an educational technology course taught in a teacher education program. Generally, this course is taken early in the teacher education program. All institutions were from the same state in the United States, and thus had similar curricula objectives to meet state teacher education standards. A review of the syllabi from each institution included within this study shows activity types consistent with the works of Harris, Mishra, and Koehler (2009). This included activities like developing technology-enhanced lesson plans, using productivity tools, and evaluating software packages for their educational affordances.

Measures

The Survey of Pre-service Teachers' Knowledge of Teaching and Technology (TKTT) theoretically measures 10 related constructs. These constructs include PK, TK, CK (which includes LCK, MCK, SSCK, and SCK), TPK, TCK, PCK, and TPACK. This section briefly describes each of these constructs. Details about the measurement properties of these constructs can be found in the results section of the article. Schmidt et al. (2009) provide a description of each construct (p. 125):

1. PK is pedagogical knowledge, which refers to the "to the methods and processes of teaching and includes knowledge in classroom management, assessment, lesson plan development, and student learning."
2. TK is technology knowledge, which refers to the "knowledge about various technologies, ranging from low-tech technologies such as pencil and paper to digital technologies such as the Internet, digital video, interactive whiteboards, and software programs."
3. CK is content knowledge, which is the "knowledge about actual subject matter that is to be learned or taught." CK was broken down into four subject areas:
    a. LCK, which refers to literacy knowledge.
    b. MCK, which refers to mathematics knowledge.
    c. SSCK, which refers to social sciences knowledge.
    d. SCK, which refers to science knowledge.
4. TPK is technological-pedagogical knowledge, which refers to "the knowledge of how various technologies can be used in teaching, and to understanding that using technology may change the way teachers teach."
5. TCK is technological-content knowledge, which refers to "the knowledge of how technology can create new representations for specific content. It suggests that teachers understand that, by using a specific technology, they can change the way learners practice and understand concepts in a specific content area."
6. PCK is pedagogical-content knowledge, which refers to "the content knowledge that deals with the teaching process. Pedagogical content knowledge is different for various content areas, as it blends both content and pedagogy with the goal being to develop better teaching practices in the content areas."
7. TPACK is technological-pedagogical-content knowledge, which refers to "the knowledge required by teachers for integrating technology into their teaching in any content area. Teachers have an intuitive understanding of the complex interplay between the three basic components of knowledge (CK, PK, TK) by teaching content using appropriate pedagogical methods and technologies."

In addition to the TKTT, we also adopted external measures of technology-related constructs to discern an external relationship with TPACK. These measures included attitudes towards technology, frequency of technology use, technology self-efficacy, and technology anxiety. The attitudes toward technology, frequency of technology use, and technology self-efficacy measures were modified from the Programme for International Student Assessment (PISA) and then slightly modified to match the language on the TKTT (PISA, 2015). The questionnaire has been rigorously analyzed to demonstrate both reliability and validity across diverse and international populations (PISA, 2015). The technology anxiety scale was inspired by an instrument designed to measure computer anxiety (Heinssen, Glass, & Knight, 1987) and was modified to reflect common and generic uses of technology.

Attitudes Toward Technology

The attitudes toward technology scale included five items that related to the use of technology in school, personal, and work settings (e.g., It is important to me to work with a computer, or Using technology helps me with my work). It was a 4-point scale ranging from 1 = *strongly disagree* to 4 = *strongly agree* without a neutral option. The scale demonstrated a high level of internal consistency reliability with Cronbach's $\alpha = .83$, which is above the social science standard of .70 (Nunnaly, 1978).

Frequency of Technology Use

The frequency of technology use scale included 10 items that required respondents to indicate the frequency with which they use a type of technology (e.g., word processing or presentation software). The response scale included five points: 1 = *never*, 2 = *less than once a month*, 3 = *between once a week and once a month*, 4 = *a few times each week*, and 5 = *almost every day*. The scale demonstrated a high level of internal consistency reliability with Cronbach's $\alpha = .82$.

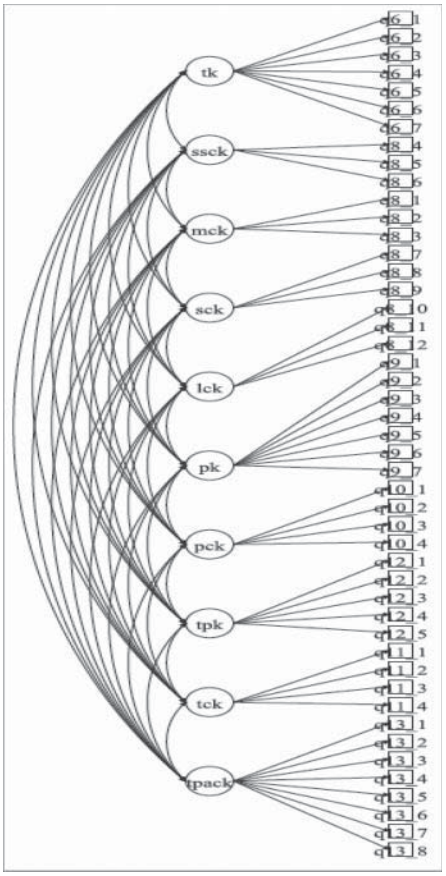

Figure 2. Visual depiction of the theoretical TPACK framework model.

Technology Self-Efficacy

The technology self-efficacy scale included 19 unique items that asked participants to rate how well they could execute a specific technology task (e.g., Scroll up and down a document on the screen, or Attach a file to an e-mail message). The response scale included four points on a continuum: 1 = *I don't know what this means*, 2 = *I know what this means but I cannot do it*, 3 = *I can do this with help from someone*, and 4 = *I can do this very well by myself*. The scale demonstrated a high degree of internal consistency reliability with Cronbach's $\alpha = .93$.

Technology Anxiety

The technology anxiety scale included 21 negatively stated items that required respondents to rate the statement (e.g., I am afraid that I will break the technology if I do something wrong, or I am afraid to use technology in front of others). The scale included a 5-point Likert scale from 1 = *strongly disagree* to 5 = *strongly agree*. The scale demonstrated a high degree of internal consistency reliability with Cronbach's $\alpha = .97$.

Procedures

After making arrangements with faculty in four different teacher education programs, we sought to collect data from students in the final weeks of a technology teacher education course. The courses were taught in both face-to-face and online settings. The faculty at the respective institutions collected data either by posting a link to the survey in their course management systems and sending reminder e-mails, or by collecting the data from their face-to-face classes on the final day of class. The link to survey was available for a 2-week period, and during this period, two reminder e-mails were sent out to the students in the online courses.

Data Analysis

The theoretical model implied by the Survey of Pre-service Teachers' Knowledge of Teaching and Technology (Schmidt et al., 2009) is a multidimensional latent factor model consisting of 10 correlated factors. Figure 2 visually displays this model with items labeled according to the survey item labels used in this study. The confirmatory factor analysis (CFA) based on this model was estimated in Mplus software (Muthén & Muthén, 2012) with maximum likelihood estimation. There were no missing data, as the Qualtrics software did not allow participants to submit the survey with incomplete response sets.

With the sample size we were able to obtain ($N = 227$), we planned to first assess if we had sufficient power to test the model in Figure 2. Specifically, we utilized the simulation results of MacCallum et al. (1999), which demonstrated that recovery of the true factor structure is a function of the communality size, the ratio of variables to factors, and the sample size. The model in Figure 2 has a variable-to-factor ratio of 47:10, and their study demonstrated that factor recovery under similar ratios was strong when communalities were high, even with sample sizes as low as $N = 60$ (MacCallum et al., 1999). Hence, we planned to assess the size of the communalities for our observed data. If they were considered large, then we planned to proceed as described above. If not, we planned to explore each of the 10 factors in Figure 2 individually with a series of 10 unidimensional CFA models. This approach to assessing power in factor analysis acknowledges that there are many observed data concerns that play a role in the ability to recover the true factor structure underlying data, falsifying some popularly held beliefs that minimum sample sizes and/or minimum variable to sample size ratios are the only determinants of power in factor analysis (Brown, 2015; de Winter, Dodou, & Wieringa, 2009; MacCallum, et al., 1990).

Fit statistics were used to judge the quality of the model fit to the data. Specifically, the root mean square error of approximation (*RMSEA*) is a goodness of fit index that penalizes for complexity of the model. Adequate fit would be indicated by $RMSEA \leq .06$ with a 90% confidence interval that has a lower and upper bound no higher than .05 and .08, respectively (Hu & Bentler, 1999). The standardized root mean square residual (*SRMR*) is a summary of residual covariances of the model, which are expected to be small if the model fits the data adequately. The benchmark was set at $SRMR \leq .08$ for acceptable fit (Hu & Bentler, 1999). The Comparative Fit Index (*CFI*) and Tucker-Lewis Index (*TLI*) assess whether the specified model fits the data better than a null model, with values greater than .95 indicating acceptable fit and values between .90 and .95 indicating marginal fit (Hu & Bentler, 1999). Chi-squared tests of fit were also used but are known to be very sensitive to sample size and very conservative indices of fit. A nonsignificant chi-squared value indicates good fit. Given adequate fit, model results such as unstandardized and standardized factor loadings and R^2 of observed variables were reviewed. Given poor fit, an exploratory factor analysis (EFA) was utilized to let the data inform the nature of the underlying dimensions and how they differed from the multidimensional factor structure implied by Schmidt et al. (2009). The EFA was completed in SPSS software, utilized principle axis factoring to mirror the use of communality variance in CFA, and implemented a promax rotation to allow for a correlated factor structure.

For each of the 10 dimensions retained from the preceding factor analytic process, external validity analysis was conducted. Specifically, the latent factor scores from each of the retained dimensions were correlated with the following external scales: attitudes toward technology, technology self-efficacy, frequency of technology use, and technology anxiety. Observed Cronbach's alphas associated with each of the scales were assessed as an indicator of internal consistency, with $\alpha < .80$ deemed a necessary observation for use of an external scale. Then, summated scores across the items of each external scale were used to represent the latent construct measured by each of said scales. The expected correlations between the latent factor scores and the external scales were such that positive correlations would be detected among attitudes toward technology, technology self-efficacy, and frequency of technology use and TK, TPK, and TPACK. We anticipated strong correlations between the measures and TK, and slightly lower measures for TPK and TPACK. We anticipated technology anxiety would have a negative correlation with TK and other technology-related constructs.

Results

To assess power for calibrating the CFA model in Figure 2, we first estimated communality variance components among the 47 items. The distribution of communality variance ranged from .50 to .88, and was negatively skewed with a median communality of .75. Only six variables had a communality estimate under .64. These communality variance estimates are considered large under both the MacCallum et al. (1999) and de Winter et al. (2009) definitions provided within their studies on power and recovery in factor analysis. This indicated that a sample size of $N = 227$ was more than sufficient to proceed with the full CFA model analysis.

The CFA based on the Schmidt et al. (2009) implied multidimensional factor structure showed poor fit to the data according to four of the five fit indices. The $RMSEA = .09$ ($CI_{90}\% = .085, .092$) was well above the recommended cutoffs for adequate fit. The $SRMR = .08$ was the one fit index result that was acceptable. The $CFI = .80$ and $TLI = .79$ were well below the benchmark for adequate fit. The chi-squared test was significant ($\chi^2(989) = 2745.80$, $p < .001$). Taken as a whole, it was clear that the multidimensional factor structure displayed in Figure 2 had poor fit to the data, and this was not due to problematic residual correlations, as indicated by the *SRMR* being the one acceptable fit index.

While EFA and CFA are often not completed on the same data set, it was deemed that this was an exceptional instance in that EFA would help further the understanding as to why the dimensionality structure shown in Figure 2 was not fitting well to the data. The EFA discussed in the method section was imposed on the observed item data after normality was assessed. While four of the 47 items showed kurtosis values greater than |3|, none showed skewness greater than |2|, and this was deemed as sufficient evidence of univariate normality, as EFA can be robust to minor violations. The sphericity assumption was met according to Bartlett's test ($\sim\chi^2(1081) = 9236.73$, $p < .001$), and the assumption of sampling adequacy was met according to a Kaiser–Meyer–Olkin index greater than .6 ($KMO = .90$).

The results of the EFA showed that there were nine factors with eigenvalues greater than one. The percent of variance explained by each factor and the standardized loadings of each item onto each factor are shown in Table 2. Standardized loadings less than |.30| are excluded from Table 2 for readability purposes. The dominant loading(s) (i.e., standardized loadings that were greater than other loadings by at least .10) for each item are highlighted in gray. The table shows that factors 1, 2, and 9 are difficult to name based on the item designations provided by Schmidt et al. (2009), and they do not align with the CFA subscale specifications.

Specifically, factor 1 is defined by a mix of both PK and PCK items, factor 2 is defined by a mix of both TPK and TPACK items, and factor 9 is defined by a mix of both PCK and TPACK items. Restated, the EFA results showed that one reason for poor fit of the CFA model was due to a lack of clear distinction in the data between the PK, PCK, TPK, and TPACK factors. The preceding interpretation of the EFA results included all items and their associated loadings greater than |.30|, without consideration of dropping items from the analysis. This approach was used because it was the most helpful in understanding why the CFA model in Figure 2 did not fit well. The results indicated that some items would need to have cross-loadings in the CFA model in order to improve fit because they loaded onto more than one factor in a similar way. These types of cross-loadings would require a renaming of factors and an acknowledgment that some of the intended constructs of measurement are difficult to isolate from some other intended constructs of measurement.

Another way to interpret the EFA results in Table 2 would be to take the simple structure approach, in which an item is eliminated from the analysis if it does not have a single loading that is at least .10 units higher than all other loadings associated with that item (Kim & Mueller, 1978). This process would result in the elimination of all PCK items and three TPACK items (i.e., Q13_2, Q13_7, and Q13_8). The interpretation would then indicate that PCK was not measured in a way distinct from the other constructs, and that only some of the TPACK items were distinct from other constructs. We have chosen to omit this interpretation from our discussion because our goal in this study was to confirm the intended structure of the full

Table 2. Factor Results From the EFA

		Factor								
		1	2	3	4	5	6	7	8	9
Schmidt subscale	Item	33.03% Variance explained	9.76% Variance explained	8.61% Variance explained	6.57% Variance explained	4.18% Variance explained	3.65% Variance explained	2.98% Variance explained	2.82% Variance explained	2.60% Variance explained
TK	Q6_1		.31	.70						
	Q6_2		.39	.76					.39	
	Q6_3			.81					.35	
	Q6_4		.40	.82					.41	
	Q6_5		.37	.74					.41	
	Q6_6		.39	.71			.33	.33		
	Q6_7		.40	.53			.37		.35	
MCK	Q8_1						.84	.44		
	Q8_2						.89	.50		
	Q8_3						.81	.44		.30
SSCK	Q8_4				.88	.32				
	Q8_5				.84	.39				
	Q8_6				.89	.40				
SCK	Q8_7			.37			.49	.86	.30	
	Q8_8			.35			.52	.87		
	Q8_9			.38			.46	.88		
LCK	Q8_10	.36	.38		.41	.90			.305	
	Q8_11	.36	.36		.39	.93				
	Q8_12	.37	.35		.38	.90				
PK	Q9_1	.77	.49			.41			.34	.45
	Q9_2	.85	.51			.38			.32	.57
	Q9_3	.86	.53			.35			.40	.52
	Q9_4	.87	.53			.32			.37	.47
	Q9_5	.89	.55			.31			.41	.50
	Q9_6	.75	.44			.35			.3	.47
	Q9_7	.70	.36							.50
PCK	Q10_1	.62	.39				.56			.62
	Q10_2	.70	.49			.63			.32	.56
	Q10_3	.58	.38				.38	.56	.32	.67
	Q10_4	.63	.43		.57	.39			.40	.54
TCK	Q11_1	.36	.43	.44			.42		.73	.47
	Q11_2	.42	.51	.32		.54			.69	.36
	Q11_3	.40	.53	.40			.31	.50	.76	.46
	Q11_4	.35	.45	.37	.47				.85	.36
TPK	Q12_1	.58	.80	.48		.38			.55	.56
	Q12_2	.57	.83	.48		.41			.58	.51
	Q12_3	.41	.73						.34	
	Q12_4	.46	.75	.37		.34			.45	.40
	Q12_5	.41	.81	.48		.37			.50	.51

(*Continued on next page*)

Table 2. *(Continued)*

		Factor								
		1	2	3	4	5	6	7	8	9
Schmidt subscale	Item	33.03% Variance explained	9.76% Variance explained	8.61% Variance explained	6.57% Variance explained	4.18% Variance explained	3.65% Variance explained	2.98% Variance explained	2.82% Variance explained	2.60% Variance explained
TPACK	Q13_1	.44	.42				.54		.36	.66
	Q13_2	.56	.61	.30		.48			.41	.68
	Q13_3	.39	.41				.33	.39	.34	.73
	Q13_4	.47	.56	.35	.43				.55	.69
	Q13_5	.57	.86	.37		.32			.48	.66
	Q13_6	.60	.81	.36		.32			.48	.70
	Q13_7	.42	.56	.43					.45	.58
	Q13_8	.568	.73	.30		.32			.43	.69

scale that has been used widely in practice without elimination of items. In other words, practitioners using this scale are not eliminating items, but rather they are interpreting summated scores on each of the 10 subscales as representative of the constructs they are hypothesized to measure. Our CFA did not support this practice, and the EFA was used only as a tool to understand the differences between the hypothesized CFA model and the observed data factor structure. We found that many of the items used in practice to represent PK, PCK, TPK, and TPACK are not distinguished from each other in the manner that theory has hypothesized and that practitioners have assumed when scoring the tool.

Under all methods of interpreting the EFA results, factors 3 through 8 align well with the six remaining subscales from Schmidt et al. (2009) and their specifications in the CFA. Factor 3 represents TK, factor 4 represents SSCK, factor 5 represents LCK, factor 6 represents MCK, factor 7 represents SCK, and factor 8 represents TCK. Presumably, the portion of the CFA related to these six factors was not responsible for the misfit problems.

Table 3 shows the factor correlation matrix estimated from the estimated sample EFA factor scores. There is a plethora of positive, statistically significant correlations between the latent factors derived from the EFA, and no statistically significant negative correlations. This matches with the Schmidt et al. (2009) intention in the full TPACK scale. However, the nonsignificant correlations between the EFA factors do not align with Schmidt et al. (2009) and may be partially responsible for some of the misfit in the CFA. In order to further understand the nonsignificant factor correlations, the nature of factors 1, 2, and 9 would first have to be determined prior to refitting a CFA and determining why some of the latent factors do not correlate as expected.

Only the subscales that were defined well in the data set were utilized for external validation purposes. PK, PCK, TPK, and TPACK were not well distinguished internally as latent factors and therefore were not related to external factors. Table 4 shows the external validation results for the TK, SSCK, LCK, MCK, SCK, and TCK scales. All subscales had positive, statistically significant correlations with attitudes toward technology, self-efficacy with technology, and frequency of technology use. The TK subscale had a negative, statistically significant correlation with technology anxiety, while all other subscales had no relationship to technology anxiety.

Table 3. EFA Factor Correlation Matrix

Factor	Factor								
	1	2	3	4	5	6	7	8	9
1	1.00								
2	.63*	1.00							
3	.23*	.53*	1.00						
4	.21*	.24*	.25*	1.00					
5	.44*	.44*	.28*	.38*	1.00				
6	.30*	.28*	.30*	−.07	.02	1.00			
7	.17*	.11	.33*	.10	−.03	.56*	1.00		
8	.47*	.63*	.53*	.30*	.31*	.23*	.19*	1.00	
9	.68*	.66*	.36*	.13	.27*	.30*	.16*	.52*	1.00

*Statistically significant at the $\alpha < .05$ level.

Discussion

Interpretation of our results must be viewed within the limitations and delimitations of our study. First, we collected data from preservice teachers

Table 4. External Validation Correlation Results

	External scale			
Factor	Attitudes toward technology	Self-efficacy with technology	Frequency of technology use	Technology anxiety
TK	.62*	.39*	.50*	−.34*
SSCK	.15*	.23*	.17*	−.01
LCK	.19*	.32*	.27*	.00
MCK	.23*	.23*	.20*	−.01
SCK	.19*	.21*	.26*	−.02
TCK	.41*	.35*	.34*	−.10

*Statistically significant at the $\alpha < .05$ level.

from four public institutions in the southeastern United States. As noted by Koh, Chai, and Tsai (2010), most studies of the TPACK framework have been assessed from samples within the United States with a few exceptions (e.g., Lee & Tsai, 2010). This facet of the literature base and our present study clearly has implications for the generalizability of our findings. Second, our data were collected in two formats, from both face-to-face courses and online courses at these four institutions, and we did not collect which course format the participants participated in at each institution. Participants may have responded differently depending on whether they were enrolled in an online version or face-to-face version of the course. Third, there are a number of potential confounding factors across the participants since the participants came from four universities in different teacher education programs with different instructions and curriculum. Finally, as with any self-report measure, the honesty and sincerity of the participants clearly influence the quality of the data collected. It should also be noted that while the participants completed the survey at the conclusion of an integrated technology course, where that course fell in their program of study varied. In light of these things, we were able to generate some important findings.

We were unable to confirm the dimensionality structure of the TPACK subscales that was implied by Schmidt et al. (2009) as a 10-factor model. Four of the five fit indices showed a poor fit for our data, leading the research team to employ an EFA to further examine the nature and structure of these data. The CFA findings from our reasonably large, multidimensional data set indicated that the dimensionality structure of the TPACK does not align with that shown in Figure 2. This poses an interesting challenge for educational technology researchers. As noted by Brantley-Dias and Ertmer (2013), TPACK contains so many different subconstructs that make it too compartmentalized for measurement purposes.

Follow-up EFA findings point to a lack of clear distinction between the PK, PCK, TPK, and TPACK subscales. These four intended subscales were shown to collapse into three unintended, and difficult-to-name, latent factors according to our data. One of our identified factors was composed of items PK and PCK, while another one of our factors includes items from both TPK and TPACK. Why do preservice teachers have difficulty in delineating among these constructs? Other researchers have reiterated the problem using diverse data sets with variations of the Schmidt et al. (2009) measurement tool. For instance, Koh, Chai, and Tsai (2010) conducted an exploratory factor analysis on a sample of 1,185 Singaporean preservice teachers and found five constructs in their model. In another study, Zelkowski, Gleason, Cox, and Bismarck (2013) focused exclusively on nearly 300 secondary mathematics preservice teachers from 15 institutions of higher education. Their analysis resulted in a seven-factor model (note that they removed LCK, SCK, and SSCK to focus only on MCK) with unusual cross-loadings. They found problems with measuring the PCK, TCK, and TPK subscales.

Fortunately, not all of the constructs on the measurement scales are performing poorly. TK, SSCK, LCK, MCK, SCK, and TCK appear to be functioning well for our data. More importantly, these constructs are correlating as we predicted with other known constructs of technology, including attitudes toward technology, technology self-efficacy, frequency of technology use, and technology anxiety. All of the constructs (attitudes toward technology, technology self-efficacy, and frequency of technology use) have significant positive relationships with TK, SSCK, LCK, MCK, SCK, and TCK with the exception of the technology anxiety measure, which has either no relationship or a significant negative correlation with TK as predicted. This external validity evidence (Cavanagh & Koehler, 2013) is an important step in characterizing the TPACK framework according to the literature.

The lack of clearly discernible boundaries between the elements of PK in the TPACK framework as measured by Schmidt et al. (2009) leads to several questions about the way preservice teachers think about or recognize TPACK and its other subconstructs. One limitation of this research is how many methods courses a preservice teacher finished before completing the survey. Unfortunately, we did not collect this information, so we cannot make any predictions. However, the educational technology course required in the state teacher education programs is generally taken early in the program of study of a preservice teacher, so we can assume that students had not enrolled in many methods courses prior to taking the survey. In conceptualizing the building of TPACK as a teacher knowledge, the resulting visualization might appear as in Figure 3. The first stage shows that PK, TK, and CK each exist completely independently of the others. In the second stage, the preservice teacher begins to develop an understanding of how each of the domains would intersect with the others and how to apply that knowledge. In the third stage, TPACK

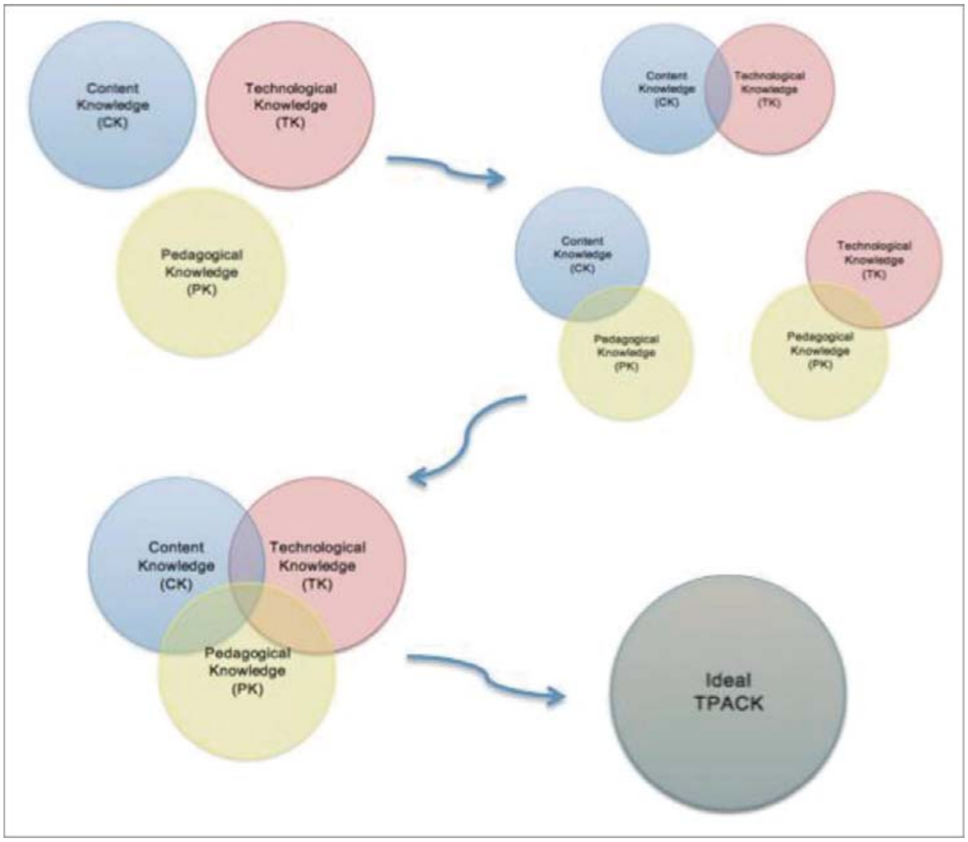

Figure 3. TPACK formulization by preservice teachers.

understanding is achieved, along with the ability to apply that knowledge. Ultimately, in the perfect teacher, TPACK knowledge domains are indistinguishable from one another in that when this teacher crafts lessons and curriculum the teacher would never consider one of the domains without the other two.

As one envisions TPACK as just described, questions arise in regard to this development and the subsequent ability of the tool to measure said development. The first question to consider is how preservice teachers develop TPACK knowledge. If preservice teachers are learning technology as part of educational theory coursework, their ability to distinguish between the PK, PCK, TPK, and TPACK subscales may be diminished. The mere fact that they are being trained to think of the pedagogy behind teaching each subject throughout various courses, while creating skewed results within the measurement tool, may be an indicator of good teacher education wherein the preservice teacher no onger retains the ability to distinguish between the domains of TPACK as they progress along the path of mastery. Therefore, the next question to ask is, if to grasp a pedagogy impacts the ability to measure TPACK, when does this happen? This is not a question that really answerable within this study, as much as it is a consideration for f uture work. Finally, does this proposed phenomenon even happen? This question may require investigation outside of preservice teachers and teacher education.

Our findings suggest that the current version of the TKTT proposed by Schmidt et al. (2009) does not measure the latent constructs implied by the 10-factor model. The poor fit indices coupled with the results from our EFA suggest that the preservice teachers in our study have problems differentiating among the many subconstructs the instrument and framework imply. Although this difference may be explained by how far into their teacher education programs our students were and the content of the educational technology course completed, these results suggest the need for further validation and possible revision of the TKTT. Future studies need to work to validate the TKTT on students in various places in their teacher education programs and then suggest possible revisions in the TKTT based on these data.

Author Notes

Dr. Albert D. Ritzhaupt is an associate professor at the University of Florida. His primary research areas focus on the design and development of technology-enhanced learning environments, and technology integration in education. His publications have appeared in multiple venues, including the Journal of Research on Technology in Education, Computers in the Schools, Journal of Computing in Higher Education, Computers & Education, Behavior Research Methods, Journal of Interactive Learning Research, Educational Technology Research and Development, Journal of Educational Computing Research, *and* Computers in Human Behavior. *Please address correspondence to Albert D. Ritzhaupt, School of Teaching and Learning, 2423 Norman Hall, PO Box 117048, Gainesville, FL 32611, USA. E-mail: aritzhaupt@gmail.com*

Dr. Anne Corinne Huggins-Manley is an assistant professor of research and evaluation methodology. Her research interests include educational measurement, applied issues in measurement, and fairness in score reporting for large-scale testing. Her primary research is on the reporting and use of disaggregated test scores for subpopulations of examinees. Through explorations of equating invariance, subpopulation item parameter drift, group dimensionality, and differential item functioning; she aims to further our understanding of the conditions that threaten validity evidence for interpreting test scores at the subpopulation level and for basing high-stakes decisions on such test scores. She also conducts research on the development of tests and surveys.

Krista Ruggles is a PhD candidate in the Curriculum, Teaching & Teacher Education program at the University of Florida with a focus in education technology. Specific areas of interest include technology integration, preservice teacher preparation, and multidisciplinary curriculum. Her current research concerns preservice teachers' understandings and implementation of asset-based lessons using technology.

Matthew Wilson is a doctoral student pursuing a degree in curriculum and instruction focusing on educational technology through the University of Florida School of Teaching & Learning. He currently teaches both undergraduate and graduate courses in educational technology for the University of Florida ProTeach Program. He has more than a decade of international teaching experience at the preschool and elementary levels. He holds an MA in early childhood and elementary education from Willamette University, and a BA in German language and literature from the University of Oregon. His research areas include technology integration in K–12 contexts and the digital divide.

References

Abbitt, J. T. (2011). An investigation of the relationship between self-efficacy beliefs about technology integration and technological pedagogical content knowledge (TPACK) among pre-service teachers. *Journal of Digital Learning in Teacher Education*, *27*(4), 134–143.

An, H., Wilder, H., & Lim, K. (2011). Preparing elementary pre-service teachers from a non-traditional student population to teach with technology. *Computers in the Schools*, *28*(2), 170–193.

Archambault, L. M., & Barnett, J. H. (2010). Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education*, *55*(4), 1656–1662.

Archambault, L., & Crippen, K. (2009). Examining TPACK among K–12 online distance educators in the United States. *Contemporary Issues in Technology and Teacher Education*, *9*(1), 71–88.

Brantley-Dias, L., & Ertmer, P. A. (2013). Goldilocks and TPACK: Is the construct "just right?" *Journal of Research on Technology in Education*, *46*(2), 103–128.

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

Cavanagh, R. F., & Koehler, M. J. (2013). A turn toward specifying validity criteria in the measurement of technological pedagogical content knowledge (TPACK). *Journal of Research on Technology in Education*, *46*(2), 129–148.

Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2010). Facilitating pre-service teachers' development of technological, pedagogical, and content knowledge (TPACK). *Educational Technology & Society*, *13*(4), 63–73.

Chai, C. S., Ling Koh, J. H., Tsai, C. C., & Lee Wee Tan, L. (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.

de Winter, J.C.F., Dodou, D., & Wieringa, P.A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *2*, 147–181.

Doering, A., Veletsianos, G., Scharber, C., & Miller, C. (2009). Using the technological, pedagogical, and content knowledge framework to design online learning environments and professional development. *Journal of Educational Computing Research*, *41*(3), 319–346.

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, *57*(3), 1953–1960.

Graham, R., Burgoyne, N., Cantrell, P., Smith, L., St Clair, L., & Harris, R. (2009). Measuring the TPACK confidence of inservice science teachers. *TechTrends*, *53*(5), 70–79.

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education*, *41*(4), 393–416.

Heinssen, R. K., Jr., Glass, C. R., & Knight, L. A. (1987). Assessing computer anxiety: Development and validation of the computer anxiety rating scale. *Computers in Human Behavior*, *3*(1), 49–59.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Keller, J. B., Bonk, C. J., & Hew, K. (2005). The TICKIT to teacher learning: Designing professional development according to situative principles. *Journal of Educational Computing Research*, *32*(4), 329–340.

Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, *32*(2), 131–152.

Koehler, M. J., Mishra, P., & Yahya, K. (2007). Tracing the development of teacher knowledge in a design seminar: Integrating content, pedagogy and technology. *Computers & Education*, *49*(3), 740–762.

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., & Divaharan, S. (2011). Developing pre-service teachers' technology integration expertise through the TPACK-developing instructional model. *Journal of Educational Computing Research*, *44*(1), 35–58.

Lee, M. H., & Tsai, C. C. (2010). Exploring teachers' perceived self efficacy and technological pedagogical content knowledge with respect to educational use of the World Wide Web. *Instructional Science* *38*, 1–21.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84–99.

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education*, *21*(5), 509–523.

Niess, M. L., van Zee, E. H., & Gillow-Wiles, H. (2010). Knowledge growth in teaching mathematics/science with spreadsheets: Moving PCK to TPACK through online professional development. *Journal of Digital Learning in Teacher Education*, *27*(2), 42–52.

Nunnaly, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Polly, D. (2011). Examining teachers' enactment of technological pedagogical and content knowledge (TPACK) in their mathematics teaching after technology integration professional development. *Journal of Computers in Mathematics and Science Teaching*, *30*(1), 37–59.

Program for International Student Assessment. (2015). *Program for International Student Assessment (PISA)*. Retrieved from http://nces.ed.gov/surveys/pisa

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK): The development and validation of an assessment instrument for pre-service teachers. *Journal of Research on Technology in Education*, *42*(2), 123–149.

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

Thompson, A. D., & Mishra, P. (2007). Editors' remarks: Breaking news: TPCK becomes TPACK! *Journal of Computing in Teacher Education*, *24*(2), 38–64.

Thompson, A., & Schmidt, D. (2010). Second-generation TPACK: Emphasis on research and practice. *Journal of Digital Learning in Teacher Education*, *26*(4), 125.

Young, J. R., Young, J. L., & Hamilton, C. (2013). The use of confidence intervals as a meta-analytic lens to summarize the effects of teacher education technology courses on pre-service teacher TPACK. *Journal of Research on Technology in Education*, *46*(2), 149–172.

Zelkowski, J., Gleason, J., Cox, D. C., & Bismarck, S. (2013). Developing and validating a reliable TPACK instrument for secondary mathematics pre-service teachers. *Journal of Research on Technology in Education*, *46*(2), 173–206.