WILLIAM & MARY
CHARTERED 1693

W&M ScholarWorks

Book Chapters

2011

Testing a TPACK-Based Technology Integration Observation Rubric

Mark J. Hofer
College of William and Mary

Neal Grandgenett
University of Nebraska at Omaha

Judi Harris
College of William and Mary

Kathy Swan
University of Kentucky

Follow this and additional works at: https://scholarworks.wm.edu/bookchapters

Part of the Educational Assessment, Evaluation, and Research Commons, and the Teacher Education and Professional Development Commons

Recommended Citation

Testing a TPACK-Based Technology Integration Observation Instrument

Mark Hofer
School of Education
College of William & Mary
Williamsburg, VA USA
mark.hofer@wm.edu

Neal Grandgenett
Teacher Education Department
University of Nebraska - Omaha
Omaha, NE USA
ngrandgenett@unomaha.edu

Judi Harris
School of Education
College of William & Mary
Williamsburg, VA USA
judi.harris@wm.edu

Kathy Swan
College of Education
University of Kentucky
Lexington, KY USA
kswan@uky.edu

Abstract: Teachers' knowledge for technology integration – conceptualized as technological pedagogical content knowledge, or TPACK (Mishra & Koehler 2006) – is difficult to discern, much less assess. Given the complexity, situatedness and interdependence of the types of knowledge represented by the TPACK construct, well-triangulated ways to assess demonstrated technology integration knowledge are needed. In 2009, three of the authors created and tested a rubric that was found to be a valid and reliable instrument to assess the TPACK evident in teachers' written lesson plans (Harris, Grandgenett & Hofer 2010). We have now also developed a TPACK-based observation rubric that testing has shown to be robust. Seven TPACK experts confirmed the rubric's construct and face validity. The instrument's interrater reliability coefficient (.802) was computed using both Intraclass Correlation and a percent score agreement (90.8%) procedure. Internal consistency (Cronbach's Alpha) was .914. Test-retest reliability (score agreement) was 93.9%. The rubric is available online at http://activitytypes.wm.edu/Assessments.

Assessing Teachers' TPACK

The specific nature of teachers' knowledge is notoriously difficult to discern, much less assess, with accuracy (Shulman 1986). It is situated, socially constructed, and highly complex (Shulman 1987; Putnam & Borko 2000). Yet if we choose to use a construct such as TPACK (Mishra & Koehler 2006) to conceptualize teachers' technology integration knowledge, we need valid and reliable strategies and instruments to assess that knowledge in the many forms in which it appears.

Recently, researchers have begun to use TPACK as a framework to explore multiple ways to understand and assess teachers' knowledge for technology integration. Data from self-reports (e.g. survey responses), interviews, and classroom observations, along with documents (e.g., preservice teachers' journal entries), and other artifacts (e.g., samples of K-12 students' work) have been analyzed – often alone, but increasingly in combination – in researchers' attempts to describe and appraise teachers' TPACK. Many recent TPACK assessments are based primarily upon survey data, whether focused on teachers' technology proficiency (e.g., Ward & Overall 2010), self-efficacy (e.g., Lee & Tsai 2010), technology adoption concerns and/or stages (e.g., Williams, Foulger & Wetzel 2010), perceptions of necessary knowledge (e.g., de Ovierira & Romero 2010; Robertshaw & Gillam 2010), or evaluations of TPACK-based learning experiences (e.g., Zhou, Zhang, Li & Zhao 2010). Several have been based upon content analyses of teachers' written reflections, such as assigned journal entries (e.g., Hechter & Phyfe 2010) and responses to instructional dilemmas (e.g., Graham, Burgoyne & Borup 2010).

Increasingly, multiple-method assessments of teachers' technology integration knowledge have been attempted in an effort to uncover and understand more of the complexity inherent in the interdependence and situatedness of the TPACK construct. Studies of experienced teachers' TPACK by Niess & Gillow-Wiles (2010) and Mueller (2010), for example, used individual interviews, observations of teaching, and a variety of self-report and self-assessment surveys, along with samples of student work (Mueller) and teachers' portfolios and online discussions (Niess & Gillow-Wiles) to help researchers to better understand the multiple dimensions of participating teachers' curriculum-based technology integration knowledge. Stoilescu and McDougall's (2010) ethnographic study of mathematics teachers' technology integration incorporated similar data types and sources, but in a more immersive and contextually-based way typical of modern anthropological research. Jaipal and Figg's (2010) multiple case study of two groups of preservice teachers triangulated data sources and types extensively, combining multiple individual interviews with each participant, multiple instructional observations and debriefings, responses to both structured and open-ended survey items, and analysis of lesson plans to determine the "characteristics of teacher knowledge for planning and implementing technologically-enhanced…instruction" (p. 3868).

TPACK Instruments

A perennial challenge posed by such rich, well-triangulated ways to assess teachers' knowledge is their dependence upon researchers' subjective interpretations of multifaceted data types. This challenge can be addressed, in part, with the use of individually validated instruments to generate some (or all) of the data considered during knowledge assessment. Unfortunately, published TPACK-based assessment instruments that have been tested for validity and reliability are still small in number and variety. Schmidt, Baran, Thompson, Koehler, Shin, & Mishra (2009) and Archambault & Crippen (2009) developed self-report instruments with multiple items keyed to each of the seven types of knowledge represented in the TPACK construct: technological (T), pedagogical (P), content (C), technological pedagogical (TP), technological content (TC), pedagogical content (PC), and technological pedagogical content knowledge (TPACK). Schmidt et al.'s survey was designed for repeated use by preservice teachers as they progress through their teacher education programs. It was also found to be reliable and valid for use at the beginning and end of shorter-duration summer courses in technology integration. Archambault and Crippen's survey instrument, designed to be used by inservice instructors, was found to be reliable and valid with a nationally representative sample of approximately 600 K-12 online teachers. In a later factor analysis of the same data set, however, Archambault and Barnett (2010) reported that the individual types of knowledge encompassed by the TPACK construct as represented in their survey instrument were "difficult to separate out…calling into question their existence in practice" (p. 1660) .

The challenges inherent in assessing teachers' knowledge accurately via self-reports—in particular, that of inexperienced teachers—are well-documented. Unfortunately, measured gains in teachers' self-assessed knowledge over time reflect their increased *confidence* regarding a particular professional development topic more than their actual increased *knowledge in practice* (Lawless & Pellegrino 2007; Schrader & Lawless 2004). Self-report data should therefore be triangulated carefully with external assessments of teachers' TPACK knowledge. Since no instrument had been developed and published by mid-2009 (to our knowledge) that supported this type of performance-based evaluation of TPACK, three of the authors decided to create and test a rubric that can be used to assess the TPACK evident in teachers' written lesson plans. Five TPACK experts confirmed the instrument's construct and face validity prior to testing. The instrument's interrater reliability was examined using both Intraclass Correlation (.857) and a percent score agreement procedure (84.1%). Internal consistency (using Cronbach's Alpha) was .911. Test-retest reliability (percent score agreement) was 87.0%. The authors have made the rubric available for noncommercial use via a Creative Commons License (Harris, Grandgenett & Hofer 2010).

Encouraged by the success of the lesson plan instrument testing, we planned the development and testing of a TPACK-based instrument that can be used to assess observed evidence of TPACK during classroom instruction. Though several validated observation instruments that focus upon technology integration have been shared with the educational technology community (e.g., Northwest Educational Technology Consortium 2004; Zambo, Wetzel, Buss & Padget 2003), none (to our knowledge) are TPACK-based. In fact, most emphasize teachers' observable technological, pedagogical, and technological pedagogical knowledge, ignoring or de-emphasizing demonstrated content/curriculum, technological content, contextual, and/or technological pedagogical content knowledge.

Instrument Testing

We drafted an observation instrument based on our Technology Integration Assessment Rubric (Harris, Grandgenett & Hofer 2010), then two authors pilot-tested it in four middle school and high school classrooms. Following each use of the rubric, the researchers conferred and made revisions to clarify items' wording and scoring. After revising the instrument several times, we sought the assistance of seven TPACK researchers from different universities to provide feedback regarding its construct and face validity. The reviewers examined the rubric, then provided focused written comments addressing seven free-response questions about the instrument. We revised some of the rubric's items, along with several aspects of its structure, according to the experts' suggestions.

We then asked 12 experienced technology-using teachers and district-based teacher educators (described in Table 1 below) in two different geographic regions of the United States to each test the reliability of the instrument by using it to assess six preservice and six inservice teachers' videorecorded lessons. The lessons were taught in elementary, middle and high school classrooms in a variety of content areas, including mathematics, language arts, science, and social studies. They were taught as part of either student teaching for preservice teachers or a professional development initiative for practicing mathematics and science teachers. The two groups of educators

("scorers") met at the researchers' two universities during either July or August of 2010 for a three-hour training and scoring practice session to learn to use the rubric in preparation for the actual scoring of each of the 12 videotaped lessons.

After the scorers used the rubric to assess each of the 12 videotaped lessons independently, they answered the same seven free-response questions about their experiences with and thoughts about the rubric to which the experts had responded earlier. We asked each educator to re-score three of the lesson recordings from a "check set" (Novak, Herman & Gearhart 1996) of the classroom videos — three recordings that the researchers had agreed represented comparatively strong, average, and weak demonstrated TPACK — via email one month after scoring them for the first time. We used these data to calculate the instrument's test-retest reliability.

Scorer	Years Taught	Content Specialty	Grade Levels Taught	Years Teaching w/ Digital Techs.	Ed Tech PD Hours: Previous 5 Years	Ed Tech Expertise Self-Assess.
A	8	Science, Math	7-8	6	400	Advanced
B	20	Math, Statistics	9-12, University	10	450	Advanced
C	10	Elementary, Science	5-6	6	300	Advanced
D	5	Elementary, Educational Technology	6, K-6 Tech.	5	500	Advanced
E	6	Elementary	K, 3, 5, 6	5	200	Advanced
F	4	Elementary reading	2	4	50	Intermediate
G	12	Special Education, Educational Technology	6-11	8	200	Advanced
H	14	Language Arts, Social Studies, Library Science	K-6, 8	12	100	Intermediate
I	30	Elementary, Educational Technology	K-12, University	20	300	Expert
J	22	English, Educational Technology	6-10, University	18	150	Advanced
K	10	Math, Gifted Education	K-2, 6-8, University	8	52	Advanced

Table 1: Study participants working at pseudononymous Midwestern and Southeastern (shaded) Universities.

Validity Analysis

The validity of the instrument was examined using two strategies that are recommended for rubric validation (cf. Arter & McTighe 2001; Moskal & Leydens 2000). Construct validity reflects how well an instrument measures a particular construct of interest, which in this study was TPACK as it is represented in observable teaching. As explained above, seven expert reviewers examined the rubric's construct validity. Face validity, or whether an instrument *appears* to informed observers to measure what it is supposed to measure, was examined using feedback from the teachers who scored the set of 12 videorecorded lessons–specifically, responses to the seven open-ended to which the expert reviewers also responded.

Construct validity was a particularly important aspect of validity to examine, since the rubric instrument was developed with TPACK as a central and unifying construct. The seven experts consulted had strong qualifications for this review process, which included extensive experience with the TPACK framework as both researchers and teacher educators. In addition, two of the reviewers authored chapters in the *Handbook of Technological Pedagogical Content Knowledge (TPCK) for Educators* (AACTE 2008), and one wrote a TPACK-based preservice textbook. These researchers were asked to gauge how well technological pedagogical knowledge (TPK), technological content knowledge (TCK) and TPACK were represented in the rubric, how well technology integration knowledge might be ascertained overall when using the rubric to evaluate an instructional episode, and what changes might be made to the rubric to help it to better reflect evidence of TPACK in observed instruction.

The experts provided a broad range of comments and suggestions in their reviews. Overall, they received the instrument quite positively. We considered all of their feedback carefully, sifting through it to identify ideas relevant to the specific TPACK-based intent and focus of our work. Many of the ideas offered were outside the scope of the instrument. For example, two reviewers suggested including a measure of the "student-centeredness" of the instruction observed, and two others suggested the addition of an item estimating student engagement. While these ideas are valid for assessing instruction in general, we realized that including them in this particular rubric would divert from our intent to develop an instrument measuring TPACK-based technology integration, irrespective of pedagogical approach observed. Based upon other feedback from the expert reviewers, we revised some of the rubric's text to clarify intent and/or to make it more uniform. For example, rather than using "learning activities" and "instructional strategies" interchangeably, we opted to use the latter throughout the rubric for consistency.

Four of the reviewers expressed concern that the rubric draft did not incorporate items addressing the instructional effectiveness and technical success in the *enacted* (rather than intended) use of technology during instruction. These concerns were particularly helpful for us to consider. The reviewers suggested that even a well-selected technology may not be *used* effectively by the teacher in the classroom setting, representing one aspect of the teacher's technological pedagogical knowledge (TPK). To help to address this concern, we added a fifth row to the rubric that addresses "Instructional Use," which we describe as "using technologies effectively for instruction." Similarly, the reviewers suggested that we add an assessment of teachers' proficiency in *operating* the technology used during the observed lesson. To include this aspect of teachers' demonstrated technological knowledge (TK), we added a sixth row entitled "Technology Logistics," which we describe as "operating technologies effectively."

The rubric's face validity was determined by reviewing the scorers' feedback on both the process of using the rubric and its perceived utility. All of the scorers' written comments supported the instrument's ability to help teacher educators to assess the quality of TPACK-based technology integration as demonstrated in observed instruction. While the scorers suggested only one small change of wording in the rubric, several noted challenges encountered when scoring lessons, and others identified limitations of using videotapes as data sources. The teachers also expressed a concern about how to differentiate assessments between two of the rows: "Instructional Strategies and Technologies" and "Instructional Use." In response, we created a scoring guide that will be described below.

Reliability Analysis

The reliability analyses for the observation rubric were conducted in July and August of 2010 with 12 teachers participating: six at Southeastern University and six at Midwestern University. The same rubric was used at each of the two locations, although the administration of the instrument at Southeastern University had preceded the administration at Midwestern University by one month, in case modifications to the rubric had been found to be necessary. Teachers/scorers at both locations were chosen purposively, based upon their experience in integrating use of digital technologies into their teaching and their diverse professional backgrounds in both content areas and grade levels. Using the data generated, reliability across both locations was calculated using four different strategies: 1) interrater reliability, computed using the Intraclass Correlation Coeficient (ICC), 2) interrater reliability, computed using a second percent score agreement procedure, 3) internal consistency within the rubric, computed using Cronbach's Alpha, and 4) test-retest reliability as represented by the percent agreement between scorings of the same videos that were completed one month apart by the same teachers. The reliability procedures used in this observation rubric study were similar to those used to validate our TPACK-based rubric for lesson plan review (Harris, Grandgenett & Hofer 2010). The procedures were selected in consultation with three expert statisticians specializing in psychometrics. These statisticians included: the director of the National Science Foundation's Center for the Assessment and Evaluation of Student Learning; an emeritus university-based measurement specialist; and a research professor working in a National Center for Research on Youth, Families, and Schools.

Each statistical procedure was selected for its particular advantages in the analysis of rubric (or similar instrument) reliability. For example, the Intraclass Correlation Coefficient flexibly examines relationships among members of a class (Field 2005; Griffin & Gonzalez 1995; McGraw & Wong 1996) and is becoming comparatively well-known in instrument validation studies. In this particular study, the educators scoring the observation videorecordings were essentially designated as a class, with rubric scores considered to be random effects, and the educators considered to be fixed effects for the ICC procedures. Percent agreement was used to further document the

extent of interrater reliability, systematically pairing scores from two different judges at a time on each video, then computing the mean percent of agreement across all teachers. Adjacent scoring was used to represent agreement, and was defined as two scores with no more than one rubric category difference. In this way, rubric scores of 3 and 4 would be considered to be in agreement, while scores of 2 and 4 would be identified as out of agreement. Percent of agreement has long been used for criterion-referenced scoring (Gronlund 1985; Litwin 2002), and it was found to be a useful way to further check interrater reliability in this study.

The rubric's internal consistency was examined using the well-established and commonly used Cronbach's Alpha procedure (Allen & Yen 2002; Cronbach, Gleser, Nanda, & Rajaratnam 1972). In this procedure, the rubric scoring data set was transposed to permit an examination of the consistency of participants' scores between each of the six rows of the rubric.

To analyze the rubric's test-retest reliability, a percent of adjacent agreement strategy was used again. The educators' scores for three of the videos were compared to their scores for the same three videos scored one month later. Each individual row's score, as well as the rubric's total scores, were compared, and an average percent agreement score was computed. As described above, the three videos selected for a second scoring process were identified as a possible "check set" of videos that were expected by the researchers to be assessed by the scorers as representing high, medium, and low levels of demonstrated TPACK. The three videos also addressed a range of content that included science, mathematics, and social studies.

Reliability Results

The scoring was conducted during the late summer of 2010, and was initiated with 12 teachers across the two institutions. However, one teacher was removed from the analysis at Midwestern University for not adhering to the timelines or process of the scoring effort. This participant exceeded the established deadline by three weeks (while all other scorers met the deadline), and needed several reminders to return the data. The participant apologized by e-mail, saying that he had been "completely swamped" with the beginning of the school year, and that he was able to do only a "quick scan" of the videos. He said that he felt that all of the teachers observed had "done very well." Upon examining the raw data contributed by this particular teacher, we noted that he had scored the videos with mostly fours on all of the rubric items, assigning no scores of 1 or 2 in any rubric row. He was the only participant to not assign any low scores for any of the lessons observed. When correlations among all scorers were run as a further check, this participant was also found to be the only one with negative correlations to other scorers. His data were thus removed from the data set, resulting in a total of 11 scorers.

To complete the Intraclass Correlation reliability procedure, the scores for each row of the rubric were recorded individually, with a total score for all six rows computed by adding the scores for each of the individual rows. Using the ICC procedure incorporated in SPSS software, the resulting statistics for the 11 judges were: Row 1 = .642, Row 2 = .661, Row 3 = .726, Row 4 = .729, Row 5 = .730, Row 6 = .834 and Total Rubric = .802. This was a comparatively strong finding for ICC, which is a statistical procedure that can produce rather conservative results for reliability computations.

The percent of agreement among the 11 scorers (after the removal of the negatively correlated participant's data described above) was also computed. This statistic is known to be less sensitive to the "direction" of how judges' scores align. Instead, it considers exclusively how "close" judges' scores are to each other. The percent agreement for the rubric scoring procedure across all scores was computed to be 90.8%, further supporting the reliability of the rubric as first calculated using ICC statistics.

The computed internal consistency of the rubric was also quite promising, calculated as .914 (Cronbach's Alpha) for the rubric as tested across the two participant groups. The rescoring of the three check set videos, which also used a percent agreement calculation, further supported the rubric's reliability. The percent agreement between the two separate scorings of the check set videos one month apart averaged 93.9% in their adjacent agreement.

Given the results of reliability testing across the 11 judges using ICC calculations, percent agreement computations, and the Cronbach's Alpha measure, we conclude that this observation instrument has comparatively strong reliability, and we feel confident in recommending it for further use. The rubric's reliability calculations,

along with its validity evaluations, suggest that we can now offer it to other researchers and educators. It has been released under a Creative Commons License, and is available on the TPACK Wiki (http://tpack.org/).

Discussion

Our testing results suggest that this is a valid and reliable instrument to use to assess enacted TPACK in observed lessons taught by either preservice or inservice teachers. We are aware of the complexities of classroom environments, however, and acknowledge that a video recording captures only what is visible in the frame, excluding important elements such as room configuration, students' facial expressions, and school and classroom climate, to name just a few. We believe, however, that the ability to analyze technology integration evidenced in videorecorded lessons provides a myriad of advantages for researchers over physical presence in the classroom. In fact, the advantages of video in the context of analyzing classroom behavior may well lie in a videorecording's ability to provide a common point of reference for viewers, focus the analysis on key observation elements, support multiple viewings as needed, and generally help to systemize the observation process (Brunvand 2010). Given these research-based advantages of videorecordings, however, it is important for researchers to remember that a video shows only the perspective of the camera's operator. What is observed may still omit important elements that are off-screen (Fishman 2003). In addition, video viewers are functioning in comparatively passive roles, without the ability for the kinds of onsite clarifications that are possible with live viewing (Barab, Hay, & Duffy 2001).

Compared to what our previous instrument (Harris, Grandgenett & Hofer 2010) examines—evidence of teachers' TPACK in written lesson plans—the data from classroom instruction that this instrument helps to generate are richer and more complex. This could lead to more or less reliable scoring. Videos of teaching may provide a better picture of enacted technology integration knowledge than a static document, and are perhaps easier to score accurately than planning documents, since many educators are more accustomed to watching people teach than to reading lesson documents in their many formats. However, there is also a greater time demand for video scoring that may negatively impact reliability. Educators who used both instruments described the richness of data available from observed teaching, as compared with what can be discerned from a written lesson plan. The greater time demands for scoring the videos were not reported to be a significant limitation upon instrument use.

Though we obtained similar validity and reliability results for both instruments, using similar statistical strategies for analysis, it was quickly apparent that guidelines and training for use of the observation instrument in particular need to be articulated clearly and carefully, to ensure that scorers are considering the TPACK-related aspects of the videotaped lessons upon which we wish to focus their attention. As with any rubric, there is always subjectivity involved in scoring, and this is true with the observation instrument presented here. We have thus developed a brief scoring guide to assist researchers with using the instrument in a consistent manner. The scoring guide is available, along with the rubric, on both the TPACK Wiki (http://tpack.org/) and the Learning Activity Types Web site (http://activitytypes.wm.edu/).

Using this instrument can help researchers and teacher educators to assess the quality of technology integration, envisioned as a teacher's TPACK-in-action. To build a more nuanced and complete understanding of enacted TPACK, however, researchers should consider using multiple data types and sources, as suggested in the introduction to this paper. We suspect that analyzing a triangulated combination of planning documents, observations, and teacher interviews (optimally, conducted both before and after observed teaching) would provide a more complete and accurate assessment of a teacher's TPACK. We invite our colleagues to use our instrument—in combination with other methods—to continue to build understanding of teachers' technology integration knowledge.

References

Allen, M.J. & Yen, W.M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

American Association of Colleges of Teacher Education (2008). *The handbook of technological pedagogical content knowledge for educators*. New York: Routledge.

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). Retrieved from http://www.citejournal.org/vol9/iss1/general/article2.cfm

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Thousand Oaks: Corwin Press, Inc.

Barab, S.A., Hay, K.E., & Duffy, T. (1999). Grounded constructions and how technology can help. *Tech Trends*, 43 (2), 15-23.

Brunvand, S. (2010). Best practices for producing video content for teacher education. *Contemporary Issues in Technology and Teacher Education*, 10 (2), 247-256.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: Wiley.

de Olvieira, J.M. & Romero, M. (2010). Student teachers' perceptions of knowledge and knowledge of perception. In J. Herrington & B. Hunter (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2010* (pp. 2351-2356). Chesapeake, VA: AACE.

Field, A.P. (2005). Intraclass correlation. In B.S. Everitt & D.C. Howell (Eds.). *Encyclopedia of statistics in the behavioral sciences* (pp. 1296-1305). Chichester, England: Wiley.

Fishman, B. (2003). Linking on-line video and curriculum to leverage community knowledge. In J. Brophy (Ed.), *Advances in research on teaching: Using video in teacher education* (Vol. 10, pp. 210-234). New York, NY: Elsevier Science.

Graham, C.R., Burgoyne, N. & Borup, J. (2010). The decision-making processes of preservice teachers as they integrate technology. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3826-3832). Chesapeake, VA: AACE.

Griffin, D., & Gonzalez, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, 118, 430-439.

Gronlund, N.E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York: McMillian.

Harris, J., Grandgenett, N., & Hofer, M. (2010). Testing a TPACK-based technology integration assessment rubric. In D. Gibson & B. Dodge (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2010* (pp. 3833-3840). Chesapeake, VA: AACE.

Hechter, R. & Phyfe, L. (2010). Using online videos in the science methods classroom as context for developing preservice teachers' awareness of the TPACK components. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3841-3848). Chesapeake, VA: AACE.

Jaipal, K. & Figg, C. (2010). Expanding the practice-based taxonomy of characteristics of TPACK. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3868-3875). Chesapeake, VA: AACE.

Lawless, K.A., & Pellegrino, J.W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research,* 77 (4), 575-614.

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), 1-21.

Litwin, M.S. (2002). How to assess and interpret survey psychometrics. *The Survey Kit Series* (Vol. 8). Thousand Oaks, CA: Sage Publications.

McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients, *Psychological Methods*, 1 (1), 30-46.

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

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7, 71-81.

Mueller, J. (2010). Observational measures of technological pedagogical content knowledge (TPACK) in the integration of laptop computers in elementary writing instruction. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3907-3910). Chesapeake, VA: AACE.

Niess, M. & Gillow-Wiles, H. (2010). Advancing K-8 mathematics and science teachers knowledge for teaching with technology through an online graduate program. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3911-3919). Chesapeake, VA: AACE.

Northwest Educational Technology Consortium (2004). Observation protocol for technology integration in the classroom (OPTIC). Portland, OR: NETC. Retrieved from http://www.netc.org/assessing/home/integration.php

Novak, J.R., Herman, J.L., & Gearhart, M. (1996). Establishing validity for performance-based assessments: An illustration for collections of student writing. *Journal of Educational Research,* 89 (4), 220-233.

Putnam, R.T. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher,* 29 (1), 4-15.

Robertshaw, M.B. & Gillam, R.B. (2010). Examining the validity of the TPACK framework from the ground up: Viewing technology integration through teachers' eyes. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3926-3931). Chesapeake, VA: AACE.

Schmidt, D., Baran, E., Thompson, A., Koehler, M.J., Shin, T, & Mishra, P. (2009, April). *Technological Pedagogical Content Knowledge (TPACK): The development and validation of an assessment instrument for preservice teachers.* Paper presented at the 2009 Annual Meeting of the American Educational Research Association, San Diego, CA. Retrieved from http://mkoehler.educ.msu.edu/unprotected_readings/TPACK_Survey/Schmidt_et_al_Survey_v1.pdf

Schrader, P. G., & Lawless, K. A. (2004). The knowledge, attitudes, and behaviors (KAB) approach: How to evaluate performance and learning in complex environments. *Performance Improvement,* 43 (9), 8–15.

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

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* 57 (1). 1-22.

Stoilescu, D. & McDougall, D. (2010). Case studies of teachers integrating computer technology in mathematics. In J. Herrington & B. Hunter (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2010* (pp. 2944-2952). Chesapeake, VA: AACE.

Ward, G. & Overall, T. (2010). Pre-service teacher technology integration: The team-taught cohort model and TPACK. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3944-3951). Chesapeake, VA: AACE.

Williams, M.K., Foulger, T. & Wetzel, K. (2010). Aspiring to reach 21st century ideals: Teacher educators' experiences in developing their TPACK. In D. Gibson & B. Dodge (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2010* (pp. 3960-3967). Chesapeake, VA: AACE.

Zambo, R., Wetzel, K., Buss, R., Padgett, H. (2003). Measuring the integration of technology through observation. In C. Crawford, N. Davis, J. Price, R. Weber, & D. Willis. (Eds.) *Technology and Teacher Education Annual* (pp. 3933 -3936). Charlottesville, VA: AACE.

Zhou, G., Zhang, Z., Li, Y. & Zhao, Z. (2010). Are secondary preservice teachers well-prepared to teach with technology? A case study of a Chinese teacher education program. In J. Herrington & B. Hunter (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2010* (pp. 2291-2300). Chesapeake, VA: AACE.

Acknowledgements