# Testing an Instrument Using Structured Interviews to Assess Experienced Teachers' TPACK

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

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

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

**Abstract:** In 2010, the authors developed, tested, and released a reliable and valid instrument that can be used to assess the quality of inexperienced teachers' TPACK by examining their detailed written lesson plans. In the current study, the same instrument was tested to see if it could be used to assess the TPACK evident in experienced teachers' planning in the form of spoken responses to semi-structured interview questions. Interrater reliability was computed using both Intraclass Correlation (.870) and a score agreement (93.6%) procedure. Internal consistency (using Cronbach's Alpha) was .895. Test-retest reliability (score agreement) was 100%. Taken together, these results demonstrate that the rubric is robust when used to analyze experienced teachers' descriptions of lessons or projects offered in response to the interview questions that appear in the Appendix.

## Assessing TPACK

During the past three years, scholarship that addresses the complex, situated, and interdependent nature of teachers' technology integration knowledge—known as "technological pedagogical content knowledge," or TPACK (Mishra & Koehler, 2006; Koehler & Mishra, 2008)—has focused increasingly upon how this knowledge can be assessed. In 2009, only five reliable and valid TPACK assessment instruments or frameworks had been published: two self-report surveys (Archambault & Crippen, 2009; Schmidt, Baran, Thompson, Koehler, Shin & Mishra, 2009), a discourse analysis framework (Koehler, Mishra & Yahya, 2007), and two triangulated performance assessments (Angeli & Valanides, 2009; Groth, Spickler, Bergner & Bardzell, 2009). By early 2012, at least eight more validated self-report survey instruments had appeared (Burgoyne, Graham, & Sudweeks, 2010; Chuang & Ho, 2011; Figg & Jaipal, 2011; Landry, 2010; Lee & Tsai, 2010; Lux, 2010; Sahin, 2011; Yurdakul, et al., 2012), along with two validated rubrics (Harris, Grandgenett & Hofer, 2010; Hofer, Grandgenett, Harris & Swan, 2011) and multiple types of TPACK-based content analyses (e.g., Graham, Borup & Smith, 2012; Hechter & Phyfe 2010; Koh & Divaharan, 2011) and verbal analyses (e.g., Mouza, 2011; Mouza & Wong, 2009) that demonstrated at least adequate levels of inter-rater reliability. Given the complexities of the TPACK construct (Cox & Graham, 2009), and the resulting challenges in its reliable and valid detection and description (cf. Koehler, Shin & Mishra, 2012), scholarship that develops and tests methods for TPACK assessment will probably continue for some time.

Our work in this area has focused upon developing and testing what Koehler et al. (2012, p. 17) term "performance assessments." These assessments "evaluate participants' TPACK by directly examining their performance on given tasks that are designed to represent complex, authentic, real-life tasks" (p. 22). Since no TPACK-based performance assessment for preservice teachers had been developed and published by mid-2009, we created and tested a rubric that can be used to assess the TPACK evident in teachers' written lesson plans (Harris, Grandgenett & Hofer, 2010). Five TPACK experts confirmed the instrument's construct and face validities prior to reliability 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%.

Given the importance of assessing both planned and enacted instruction, we then developed and tested another TPACK-based rubric that can be used to assess observed evidence of TPACK during classroom instruction (Hofer, Grandgenett, Harris & Swan, 2011). Seven TPACK experts confirmed this observation instrument's construct and face validities. Its interrater reliability coefficient was computed using the same methods applied to the lesson plan rubric, with both Intraclass Correlation (.802) and percent score agreement (90.8%) procedures. Internal consistency (Cronbach's Alpha) for the observation rubric was .914. Test-retest reliability (score agreement) was 93.9%.

## Experienced vs. Inexperienced Teachers' Planning

Our TPACK-based observation instrument (Hofer, et al., 2011) was tested using unedited classroom videos of equal numbers of both experienced and inexperienced teachers teaching. Considering this, and given the reliability and validity results summarized above, the observation rubric is sufficiently robust to be used to observe either preservice or inservice teachers. Our previous instrument (Harris, et al., 2010), however, was tested only with inexperienced teachers' lesson plans. Therefore, it was demonstrated to be a reliable and valid tool to use to assess only preservice teachers' written instructional plans. In the current study, we sought a similarly succinct, yet robust measure of experienced teachers' instructional planning with reference to the quality of their technology integration knowledge, or TPACK.

Studies of experienced teachers' lesson planning show it to be quite different from that of inexperienced teachers (Leinhardt, 1993). Inservice teachers' written plans rarely encompass everything that the teacher expects to happen during the planned instructional time, and they are not often written in a linear sequence from learning goals to learning activities to assessments (Clark & Peterson, 1986). They tend to focus upon guiding students' thinking moreso than inexperienced teachers' plans do, anticipating difficulties that students might have with the content to be taught. Experienced teachers also tend to be able to think simultaneously about their own actions, while also attending to and predicting their students' probable misconceptions and actions. Novice teachers generally do not plan or teach "in stereo" in this way, as inservice teachers do, and their actions during teaching don't always address the learning goals of the lesson completely (Leinhardt, 1993). Many experienced teachers can address the content of a lesson while meeting planned instructional objectives, connecting the content taught to larger issues, and anticipating students' probable confusions and difficulties. Inexperienced teachers tend to have much more limited knowledge of the nature of student learning, and experience difficulty in finding ways other than those that reflect their own thinking patterns to explain concepts to their students (Livingston & Borko, 1990).

Inservice teachers' written lesson plans tend to comprise brief notes only (Leinhardt, 1993), though their authors are able to explain at length the content foci, assessment strategies, targeted student thinking, alternative explanations, and "Plan B" learning activities that those limited written notes represent. Given the brevity and idiosyncrasy of experienced teachers' written planning documents, we realized that we could not assess their lesson plans in the same way that we assessed inexperienced teachers' planning artifacts. Instead, we devised a 20" – 30" semi-structured lesson interview protocol (see Appendix) that we used with volunteer inservice teachers to record essential information about their technology integrated lesson plans. These audiorecordings then became the data to which our "scorers" listened. For each interview, the scorers completed a copy of the Technology Integration Assessment Rubric (see Appendix), using it to assess the quality of the interviewed teachers' TPACK. In this way, we tested the existing rubric for reliability and validity when it was used to assess the quality of TPACK represented in experienced teachers' interactive descriptions of particular technology-infused lessons or projects.

## Instrument Testing Procedures

Twelve experienced technology-using teachers (described in Table 1 below) and district-based teacher educators in two different geographic regions of the United States tested the reliability of the lesson plan instrument when it was used to individually assess 12 inservice teachers' audiorecorded interviews about self-selected, technology-infused lessons that they planned and taught. These two groups of scorers met at two different universities during either July or August of 2011 for approximately 3 hours to learn to use the rubric with two sample lesson plan interviews, then applied it within the following two weeks to evaluate of each of the audiorecorded 12 lesson interviews. The planning interviews addressed varying content areas and grade levels.

After the scorers used the existing rubric to individually assess each of the audiorecorded lesson interviews, they answered seven free-response questions that requested feedback about using the rubric with this type of data. We also asked each scorer to re-score three assigned lesson interviews one month after scoring them for the first time, and used these data to calculate the test-retest reliability of the instrument.

| Scorer | Years Taught | Content Specialty | Grade Levels Taught | Years Teaching w/ Digital Techs. | Ed Tech PD Hours: Prev. 5 Years | Ed Tech Expertise Self-Assess. |
|--------|--------------|-------------------|---------------------|----------------------------------|--------------------------------|-------------------------------|
| A | 20 | Social Studies | 9-12 | 20 | 220 | Advanced |
| B | 11 | Elementary gifted learners | 3, 5, 6, 8 | 5 | 65 | Advanced |
| C | 12 | Elementary; Science | 3-6 | 12 | 70 | Advanced |
| D | 39 | Math | K-12 | 19 | 300 | Intermediate |
| E | 5 | Physics | 9-12 | 5 | 35 | Intermediate |
| F | 11 | Technology Integration | K-8 | 6 | 150 | Advanced |
| I | 4 | Elementary, Reading | 2 | 4 | 100 | Intermediate |
| J | 14 | Special Education | 5-12 | 9 | 200 | Advanced |
| K | 12 | English | 10-12 | 11 | 300 | Advanced |
| L | 11 | Math, Technology Integration | K-12 | 11 | 520 | Advanced |
| M | 30 | Gifted Ed., Technology Integration | K-12 & college | 25 | 120 | Advanced |
| N | 9 | Math, Gifted Ed. | K-1, 7-8 | 7 | 90 | Advanced |

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

## Validity Analysis

The construct and face validities of the instrument were examined when the instrument was first tested with preservice teachers' lesson plans (Harris, Grandgenett & Hofer, 2010). We used 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 educational lesson plans. As explained above, construct validity was examined in this study using expert reviews. Face validity, or whether an instrument appears to informed observers to measure what it is designed to measure, was examined using the experienced teachers' (scorers') responses to the seven-item survey, also described above.

Construct validity was a particularly important aspect of this rubric for us to test, since it was developed with TPACK as a central and unifying construct. The six experts consulted when the rubric was first developed and tested 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 had recently released a TPACK-based preservice textbook. The researchers were asked to gauge how well TPK, TCK and TPACK were represented in the rubric, how well technology integration knowledge might be ascertained overall when using the rubric to evaluate a lesson/project plan, and what changes might be made to the rubric to help it to better reflect evidence of TPACK in teachers' planning documents. The rubric's construct validity was supported strongly by comments from five of the six expert reviewers. The sixth expert did not agree that the quality of technology integration (and therefore teachers' TPACK) could be ascertained overall for any instructional plan. Instead, this reviewer suggested creating specific questions to be answered about the appropriateness of technology use in different aspects of an instructional plan, such as the communication of content, the instruction itself, and the assessment.

The rubric's face validity was determined by analyzing the scorers' feedback on both the process of using

the rubric and its perceived utility. All of the scorers' written comments during each of the two rubric tests (in 2009 and 2011) supported its ability to help teacher educators to assess the quality of TPACK-based technology integration inferred from lesson plans/interviews. Some also offered suggestions for minor changes to the wording in some of the rubric's cells, several of which were used to create the version of the rubric that appears in the Appendix.

## Reliability Analysis

The reliability analyses for the rubric when it was used to assess audio interviews were conducted in July and August of 2011 with 12 teachers participating: six at Southeastern University and six at Midwestern University. The same rubric was used at each of the two locations. Scorers at both locations were chosen purposely, 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 Coefficient (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 examined one month apart by the same teachers. The reliability procedures used for this study were similar to those used to validate the rubric for written lesson plan review (Harris, Grandgenett & Hofer 2010) and, in an expanded form, for the review of video observations of classes,(Hofer, Grandgenett, Harris & Swan, 2011). The statistical procedures were selected in consultation with three expert statisticians specializing in psychometrics.

Similar to our previous studies, the statistical procedures for the review of the rubric's reliability for audio interviews were selected based on each procedure's particular advantages for examining rubric reliability (or for that of similar scoring instruments). 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. It is now a scale analysis option in *SPSS* software. In this particular study, the educators scoring the audio interviews were essentially designated as a class, with rubric scores considered to be random effects, and the educators themselves representing fixed effects for the ICC calculations. 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 judges. Adjacent scoring was used to represent this scorer agreement, and was defined as two scores with no more than one rubric category of 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); it was a useful way to further check the interrater reliability of the rubric in this study.

The rubric's internal consistency in assessing the TPACK evident in audio interviews was again 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 within the SPSS data file to permit an examination of the consistency of participants' scores between each of the four 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 audio files were compared to their scores for the same three audio files scored one month earlier. Each individual row's score, as well as the rubric's total scores, were compared, and an average percent agreement score was computed. The three audio interviews selected for a second scoring process were identified as a possible "check set" of audio files that the researchers expected to be scored as representing high, medium, and low levels of demonstrated TPACK. The three recordings also represented a range of content that included elementary science, high school mathematics, and middle school foreign language.

Finally, to provide some context on the scorers' own perceptions of expertise to do such scorings adequately, the scorers assessed their expertise levels at both the time of the initial scoring and when rescoring interviews to determine if their self-perceptions of technology expertise had changed from one scoring to the next. The scorers' self-assessments confirmed their perceptions of adequate expertise. The 12 scorers all ranked themselves similarly from the first scoring to the second, with 9 scorers ranking themselves as "advanced" on the

first scoring, and three ranking themselves as "intermediate." At the time of the rescoring one month later, one of the scorers increased their ranking from intermediate to "expert," while all others scorers retained their original self-perceived levels of expertise within the intermediate and advanced categories.

### *Reliability Results*

To complete the Intraclass Correlation reliability calculation, the scores for each row of the rubric were recorded individually, with a total score for all four rows computed by adding the scores for each of the individual rows. Using the ICC procedure incorporated into SPSS software, the resulting statistics for the 12 scorers were: Row 1 = .651, Row 2 = .814, Row 3 = .681, Row 4 = .853, and Total Rubric = .818. This was a comparatively strong finding for ICC, which is a statistical procedure that can produce rather conservative results for reliability computations. However, upon further examination of the correlations among individual scorers, it was noted that one scorer was negatively correlated with all other scorers on all row scores, as well in the total scores. When that single scorer was removed, the ICC coefficients increased significantly, with Row 1 = .750, Row 2 = .850, Row 3 = .771, Row 4 = .886, and Total Rubric = .870. Upon reflection on the background this single scorer, it was determined that he was relatively unique among the set of judges in his perspective (with lower scores for many of the planning interviews), and had just assumed a full-time administrative post in his local school district. Thus, his "unique administrative perspective" on the audio lessons was different enough to warrant the removal of his scores from the data set.

The percent of agreement among the 12 scorers 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 scorers was computed to be 91.7%, further supporting the reliability of the rubric as first calculated using ICC statistics. When the negatively correlated scorer mentioned previously was removed, than the percent of adjacent agreement between scorers increased slightly to 93.6.

The computed internal consistency of the rubric was also quite promising, calculated as .895 (Cronbach's Alpha) for the rubric as tested across the 12 scorers. 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 attained 100% adjacent agreement, showing strong consistency in the separate scorings of the check set videos for all scorers.

## Conclusion

Given the results of reliability testing with 12 scorers using ICC calculations, percent agreement computations, and the Cronbach's Alpha measure, we conclude that this instrument has comparatively strong reliability for examining audio interviews describing lessons and projects in which educational technologies are incorporated, 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 to assess the TPACK evident in structured interviews done with experienced teachers. It has been released under a Creative Commons License, and is available on the Learning Activity Types Web site (http://activitytypes.wm.edu/).

We are pleased to place the interview prompts used with this instrument into the public domain, also via a Creative Commons (attribution, noncommercial, no derivatives) license, and encourage consideration of their use for both research and professional development. Given the increasing variety of tested TPACK-based instruments currently available, it is now possible to more accurately and comprehensively assess teachers' TPACK in authentic ways. We hope that the work described in this chapter will support that ongoing effort within future well-triangulated studies of teachers' TPACK.

## References

Abbitt, J. (2011). Measuring technological pedagogical content knowledge in preservice teacher education: A review of current methods and instruments. *Journal of Research on Technology in Education, 43*(4), 281–300.

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

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

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers and Education, 52*(1), 154-168.

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.

Britten, J.S., & Cassady, J.C. (2005). The Technology Integration Assessment Instrument: Understanding planned use of technology by classroom teachers. *Computers in the Schools, 22*(3), 49-61.

Burgoyne, N., Graham, C.R. & Sudweeks, R. (2010). The validation of an instrument measuring TPACK. In D. Gibson & B. Dodge (Eds.), *Proceedings of Society for Information Technology & Teacher Education international conference 2010* (pp. 3787-3794). Chesapeake, VA: AACE. Retrieved from http://www.editlib.org/p/33971

Chuang, H-H, & Ho, C-J. (2011). An investigation of early childhood teachers' technological pedagogical content knowledge (TPACK) in Taiwan. *Journal of Kirsehir Education Faculty, 12*(2), 99-117. Retrieved from http://www.doaj.org/doaj?func=abstract&id=782294&recNo=6&toc=1&uiLanguage=en

Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York: Macmillan.

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.

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

Figg, C. & Jaipal, K. (2011). Developing a survey from a taxonomy of characteristics for TK, TCK, and TPK to assess teacher candidates' knowledge of teaching with technology. In M. Koehler & P. Mishra (Eds.), *Proceedings of Society for Information Technology & Teacher Education international conference 2011* (pp. 4330-4339). Chesapeake, VA: AACE. Retrieved from http://www.editlib.org/p/37012

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

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.

Groth, R., Spickler, D., Bergner, J., & Bardzell, M. (2009). A qualitative approach to assessing technological pedagogical content knowledge. *Contemporary Issues in Technology and Teacher Education*, *9*(4). Retrieved December 5, 2009, from http://www.citejournal.org/vol9/iss4/mathematics/article1.cfm

Hammond, T. C., Alexander, R. C., & Bodzin, A. M. (2012). Assessment in authentic environments: Designing instruments and reporting results from classroom-based TPACK research. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches* (pp. 32-57). Hershey, PA: IGI Global. doi: 10.4018/978-1-60960-750-0.ch003

Harris, J., Grandgenett, N., & Hofer, M. (2010). Testing a TPACK-based technology integration assessment rubric. In C. D. Maddux (Ed.), *Research highlights in technology and teacher education 2010* (pp. 323-331). Chesapeake, VA: Society for Information Technology & Teacher Education (SITE).

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.

Hofer, M., Grandgenett, N., Harris, J., & Swan, K. (2011). Testing a TPACK-based technology integration observation instrument. In C. D. Maddux  (Ed.), *Research highlights in technology and teacher education 2011* (pp. 39-46). Chesapeake, VA: Society for Information Technology & Teacher Education (SITE).

Koehler, M.J., & Mishra, P. (2008). Introducing TPACK. In AACTE Committee on Innovation & Technology (Eds.). *Handbook of technological pedagogical content knowledge for educators* (pp. 3-29). New York, NY: Routledge.

Koehler, M.J., & Mishra, P. (2005). Teachers learning technology by design. *Journal of Computing in Teacher Education, 21*(3), 94-102.

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

Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches* (pp. 16-31). Hershey, PA: IGI Global. doi: 10.4018/978-1-60960-750-0.ch002

Koh, J. H. L., & 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. doi: 10.2190/EC.44.1.c

Landry, G. A. (2010). *Creating and validating an instrument to measure middle school mathematics teachers' technological pedagogical content knowledge (TPACK)* (Doctoral dissertation, University of Tennessee - Knoxville). Retrieved from http://trace.tennessee.edu/utk_graddiss/720

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.

Leinhardt, G. (1993). On teaching. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 4, pp. 1-54). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Livingston, C., & Borko, H. (1990). High school mathematics review lessons: Expert-novice distinctions. *Journal for Research in Mathematics Education, 21*(5), 372-387.

Lux, N. J. (2010). *Assessing technological pedagogical content knowledge* (Doctoral dissertation)**.** Retrieved from ProQuest Dissertation and Theses. (AAT 3430401)

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.

Mouza, C. (2011). Promoting urban teachers' understanding of technology, content, and pedagogy in the context of case development. *Journal of Research on Technology in Education, 44*(1), 1–29.

Mouza, C. & Wong, W. (2009). Studying classroom practice: Case development for professional learning in technology integration. *Journal of Technology and Teacher Education, 17*(2), 175-202.

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

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

Yurdakul, I. K., Odabasi, H. F., Kilicer, K., Coklar, A. N., Birinci, G., & Kurt, A. A. (2012). The development, validity and reliability of TPACK-deep: A technological pedagogical content knowledge scale. *Computers & Education 58*(3), 964–977. doi: 10.1016/j.compedu.2011.10.012

## Appendix: Interview Protocol and Assessment Rubric

**Interview Protocol**

**LESSON DESCRIPTION:**
Describe the content and/or process topic(s) for the lesson.
Describe the student learning goals/objectives addressed in the lesson. (These will not necessarily be state or national standards. Participants should describe these in their own words.)
Describe your students (e.g. grade level, and specific learning needs/preferences).
Walk me through the lesson/project as it unfolded in the classroom.
What educational technologies (digital and non-digital) did you use and how did you and/or your students use them?
Describe any contextual information (e.g. access to a computer lab, materials and resources available; particular departmental/school-wide initiatives) that influenced the design or implementation of the lesson/project.

**TPACK-SPECIFIC QUESTIONS:**
How and why do the particular technologies used in this lesson/project "fit" the content/process goals?
How and why do the particular technologies used in this lesson/project "fit" the instructional strategies you used?
How and why do the learning goals, instructional strategies, and technologies used all fit together in this lesson/project?

**Assessment Rubric**

| Criteria | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| **Curriculum Goals & Technologies** <br><br>(Curriculum-based technology use) | Technologies selected for use in the instructional plan are <u>strongly aligned</u> with one or more curriculum goals. | Technologies selected for use in the instructional plan are <u>aligned</u> with one or more curriculum goals. | Technologies selected for use in the instructional plan are <u>partially aligned</u> with one or more curriculum goals. | Technologies selected for use in the instructional plan are <u>not aligned</u> with any curriculum goals. |
| **Instructional Strategies & Technologies** <br><br>(Using technology in teaching/learning) | Technology use <u>optimally supports</u> instructional strategies. | Technology use <u>supports</u> instructional strategies. | Technology use <u>minimally supports</u> instructional strategies. | Technology use <u>does not support</u> instructional strategies. |
| **Technology Selection(s)** <br><br>(Compatibility with curriculum goals & instructional strategies) | Technology selection(s) are <u>exemplary</u>, given curriculum goal(s) and instructional strategies. | Technology selection(s) are <u>appropriate, but not exemplary</u>, given curriculum goal(s) and instructional strategies. | Technology selection(s) are <u>marginally appropriate</u>, given curriculum goal(s) and instructional strategies. | Technology selection(s) are <u>inappropriate</u>, given curriculum goal(s) and instructional strategies. |
| **"Fit"** <br><br>(Content, pedagogy and technology together) | Content, instructional strategies and technology <u>fit together strongly</u> within the instructional plan. | Content, instructional strategies and technology <u>fit together</u> within the instructional plan. | Content, instructional strategies and technology <u>fit together somewhat</u> within the instructional plan. | Content, instructional strategies and technology <u>do not fit together</u> within the instructional plan. |