



The first principles of instruction: an examination of their impact on preservice teachers' TPACK

Jacob A. Hall¹ · Jing Lei² · Qiu Wang²

Accepted: 29 October 2020 / Published online: 12 November 2020
© Association for Educational Communications and Technology 2020

Abstract

The flipped approach has been widely adopted in higher education, yet its theoretical framework and use in teacher preparation courses have been limited. To address these gaps, this study examined the impact of the First Principles of Instruction when applied to designing face-to-face and flipped technology integration courses. Participants were 32 preservice teachers enrolled during the 2017 spring and fall semesters. Employing a 3-way mixed factorial research design, we measured participants' technological, pedagogical, content knowledge (TPACK) outcomes in each group and compared the outcomes between the face-to-face and flipped groups. In both groups, preservice teachers' self-perceptions and application of TPACK statistically significantly increased. The magnitude of the TPACK application results (F2F $p < .001$, $d = 1.17$; Flipped^a $p < .001$, $d = 1.97$) strongly demonstrates the First Principles' potential to frame effective course design. Further analyses revealed no statistically significant differences between groups' TPACK outcomes. These non-significant differences suggest the First Principles of Instruction may be equally effective for designing flipped and face-to-face courses. We conclude the article by discussing implications for course design and detailing considerations for future research on flipped approaches.

Keywords First principles of instruction · TPACK · Flipped model · Teacher education · Factorial design

✉ Jacob A. Hall
jacob.hall@cortland.edu

Jing Lei
jlei@syr.edu

Qiu Wang
wangqiu@syr.edu

¹ Childhood/Early Childhood Education Department, SUNY Cortland, Education Building, Room 1247, P.O. Box 2000, Cortland, NY 13045, USA

² School of Education, Syracuse University, 252 Huntington Hall, Syracuse, NY 13244, USA

Introduction

Digital technologies have introduced extensive and rapid change (Beloit College 2016; Coles et al. 2006). Even the comforting ebb and flow of digital change may be disappearing as the era of predictable technological growth has been forecast to end (“After Moore’s Law,” 2016). Amidst technology’s vicissitudes, ubiquity, and unrealized promises, instructional designers and teachers situate their work. Environments have been blended, users immersed, reality augmented, and learners characterized as app-dependent digital natives (Gardner and Davis 2014; L. Johnson et al. 2016). Yet Merrill claims that fundamental means of learning have not changed and proposed an instructional model built on the premise that what promoted learning in the past will still be effective today and into the future (2012).

Merrill developed the First Principles of Instruction (FPI) through a synthesis of instructional design models and theories (Merrill 2002). The principles and their corollaries stem from what Merrill claimed to be the overlapping components of all theories and models he analyzed. He posits that instruction will be more effective, efficient, and engaging in proportion to its implementation of the FPI and regardless of context, approach, or audience. Although limited empirical research has evaluated Merrill’s claims, initial findings have revealed positive impacts on mastery of course objectives (Frick et al. 2010), deep cognitive strategy use (Lee and Koszalka 2016) and domain-specific critical thinking skills (Tiruneh et al. 2016). Furthermore, faculty designing courses with the FPI observed increased student ratings on course evaluations and positive feedback related to engagement and the relevance of course activities (Cheung and Hew 2015; Hoffman 2014). Given these initial results, there is reason to continue examining the efficacy of the FPI in additional disciplines and applied to varied approaches.

Therefore, this study investigated the impact of the FPI on preservice teachers’ self-perceptions and application of technological, pedagogical, content knowledge (TPACK) in both face-to-face (F2F) and flipped course approaches. It sought to answer the following research questions: 1) how do the FPI, when applied to a face-to-face course, impact preservice teachers’ self-perceptions and application of TPACK, 2) how do the FPI, when applied to a flipped course, impact preservice teachers’ self-perceptions and application of TPACK, and 3) does the impact of the FPI on preservice teachers’ self-perceptions and application of TPACK differ when applied to face-to-face and flipped approaches?

We will begin the article with an overview of the FPI’s theoretical foundations and a review of research on the FPI and its application to flipped course design. Afterward, we will synthesize research on TPACK and define the TPACK constructs in this study. Next, we will detail this study’s design and the instruments used to measure the TPACK-based learning outcomes. Results will then be shared for each research question along with an overview of the descriptive statistics for both TPACK measures. We will then discuss the results and offer suggestions for future research and instructional design.

Literature review

Theoretical Foundation of the FPI

The founding premise of the first principles is that they are applicable regardless of context or instructional program and necessary for effective, efficient, and engaging instruction. Merrill's goal was to identify principles of instruction that were fundamental to most instructional design theories and models. According to Merrill (2002), a principle is a "relationship that is always true under appropriate conditions regardless of program or practice" (pg. 43). Briefly stated, the five first principles that resulted from his synthesis are that learning is promoted when: (1) learners solve real world problems, (2) prior knowledge is activated to serve as a foundation for new knowledge, and new knowledge is (3) demonstrated to learners, (4) applied by learners, and (5) integrated into their lives.

The FPI are indirectly supported by a large body of instructional design theories and models from which Merrill drew these fundamental principles (Andre 1997; Jonassen 1999; Merriënboer et al. 2002; Nelson 1999; Schank et al. 1999). Additionally, the FPI have been regarded as foundational knowledge for the training of instructional designers (Donaldson 2017), applied to empirical research in various settings (Lee and Koszalka 2016; Tiruneh et al. 2016), and used to conceptually frame instruction (Gardner and Belland 2012; Hall and Lei 2020; Nelson 2015). Their underlying premises, however, have also been questioned. Reigeluth (2013) proposed an alternative view to Merrill's premise for universally applicable principles when he introduced the term *situationalities* and the potential need for instructional principles sensitive to contextual variations. Although the FPI offer a robust, general summary of how quality instruction can be produced, Reigeluth (2013) contended they may not address the plethora of nuances instructional situations often generate.

The generality of the FPI, however, affords them the flexibility of being applied by instructors and designers from many theoretical perspectives (Merrill 2012). No specific philosophy or learning theory need be adopted to implement these principles. Regarding theories of learning, it has been said that theoretical eclecticism may be a strength of the field (Ertmer and Newby 1993). Although some express concern for this remixing of instructional strategies advocated by theoretical eclecticism (Bednar et al. 1991), proponents argue that the instructional strategies derived from *learning theory A* are often the same as those derived from *learning theory B*. It is the theoretical rationale for each instructional move that differs. In practice, the implementation of the instruction may not be distinguishably different by supporters of one theory versus the other. However, research on the implementation of the FPI's in practice, thus framed by various theoretical rationales, is yet in its early stages.

Research on FPI and Flipped Course Design

Although there is compelling indirect support for the FPI, direct empirical evidence has been limited (Lee and Koszalka 2016; Tiruneh et al. 2016). Scholars have specifically identified the need for researching the FPI's impact in varied contexts such as additional disciplines, emerging learning environments (e.g., online, blended, flipped), and informal

environments such as museum education (Cheung and Hew 2015; Nelson 2015). The initial empirical findings have substantiated Merrill's claims regarding the FPI's efficacy (Frick et al. 2010; Lo and Hew 2017), laid a foundation for future research, and evidenced limitations.

This study's design addressed three specific limitations noted in the literature that investigated the application of Merrill's FPI in blended environments. First, the varying operationalization of blended approaches may contribute to low external validity and inhibit the synthesis of empirical findings (Bellini and Rumrill 1999). For example, in Hoffman's (2014) study of a FPI-based flipped course design, the flipped approach is operationalized as a shift in curricular sequence with no reference to how the course mixed instructor-transmitted and technology-mediated methods (Hall 2018; Margulieux et al. 2016). Hoffman described the approach as "the 'flipped' sequencing of putting analysis, verification, and conclusions as the starting point for learning (2014, pg. 57)." Although Hoffman's (2014) initial definitions of flipped aligned with previous literature on blended learning, the articulation of the FPI-based "flipped" course design was inconsistent with these definitions. In contrast, this study seeks to articulate an implementation of flipped that is consistent with a broader understanding of this approach.

Secondly, there is wide variance in how the FPI's principles are applied and the degree to which their corollaries are acknowledged or detailed within a study. In some studies, it is unclear how all principles were applied to the course design (Hoffman 2014). At other times, the principles are introduced, but it is unclear how the design instantiated them. In a study conducted by Cheung and Hew (2015), their multimedia design problem is not described, nor are its underlying tasks, operations and actions. A later study (Lo and Hew 2017), however, describes the problem in detail and discusses how elements of the problem are introduced, sequenced, engaged. As theorists have many ways of thinking about problems (Merrill 2002), this study attempted to build on Lo and Hew's (2017) precedent by clearly stipulating the elements of the problem and other implementations of the instructional principles and corollaries in the course design.

Lastly, this study builds upon previous empirical work by incorporating an objective measure of student learning with student self-perceptions. Past studies have effectively utilized student perceptions of Merrill's FPI and correlated these perceptions with various outcomes such as engagement or motivation (Lee and Koszalka 2016). Frick et al., (2009) explicitly focus on student perceptions of instructional quality rather than the design process. Their incorporation of the FPI in the Teaching and Learning Quality (TALQ) scales and the resulting conclusions have been regarded as perhaps the most robust empirical support for the FPI. One finding was that students who reported engagement in the course and the presence of the FPI were "three to five times more likely to agree or strongly agree that they learned a lot and were satisfied with courses...[and] nine times more likely to report mastery course objectives (Frick et al. 2009, p. 713)." A follow-up study built upon these findings with a much larger sample, recruitment of entire classes of students, and the addition of instructor's ratings of students' mastery. Findings from this study further supported that perceived application of FPI are predictive of perceived engagement with course tasks—when combined, these are predictive of student learning (Frick et al. 2010). While these results affirm FPI's potential for improving instruction, the studies did not aim to explore the design process or incorporate objective measures of student learning. Studies that support causal inferences, detail prescriptive design processes, or more richly describe participants' learning experiences are needed.

Several studies have attempted to measure learning outcomes resulting from FPI-based course designs. Specific to flipped approaches, many have used single group pre-post

designs (Hall 2018; Lo and Hew 2017). The current study was preceded by a single group pre-post design that examined the how preservice teachers' understandings and application of TPACK were influenced by a flipped course designed according to the FPI (Hall 2018). Most notably, participants significantly improved their application of TPACK when designing technology-integrated lessons, and they increased their self-perceptions of PK. While participants in Lo and Hew's (2017) study also demonstrated statistically significant learning gains, the addition of control groups could offer greater clarity regarding the impact of the intervention (Creswell 2012). In both studies, the question remains whether the significant learning gains were related to the FPI, the flipped approach, or an interaction effect. To begin addressing this limitation, the current study tested the FPI with multiple measures of TPACK-based outcomes within a flipped course and a comparably designed face-to-face (F2F) course.

TPACK in teacher preparation

While similar ideas were discussed in literature prior to Koehler and Mishra (Angeli et al. 2016), TPACK became especially popular for developing teachers' technology integration knowledge and skills after Mishra and Koehler's first article detailing TPCK (2006). Soon thereafter, it was renamed TPACK as it was meant to comprise the Total PACKAGE of knowledge needed by teachers to effectively integrate technology within a given context (Koehler and Mishra 2009). Its use has ranged from reimagining practice in K-12 and higher education (Chai et al. 2013) to reforming teacher preparation programs' methods for developing preservice teachers. Teacher educators frame learning outcomes in technology integration courses with TPACK as it extends Shulman's (1986) seminal integration of content and pedagogical knowledge to seven domains of knowledge: three distinct domains (TK, PK, and CK) and four domains (TPK, TCK, PCK, and TPACK) formed from the overlap.

Schmidt et al., developers of the survey instrument used in this study, operationally defined TPACK as "the knowledge required by teachers for integrating technology in 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 (2009, p. 125)." Abbreviated definitions for each subdomain, drawn from Schmidt et al.'s (2009) definitions for TPACK constructs, are presented below:

TK is the knowledge of how to use various digital and non-digital technologies.

CK is the knowledge of the subject or discipline without reference of how to teach it.

PK is the knowledge of instructional practices and student learning, assessment and classroom management and is independent of content and technology.

PCK is the knowledge of how to implement and adapt instructional practices in ways most appropriate for specific content areas.

TPK is the knowledge of technologies and their affordances that would enhance teaching approaches and learning environments with no consideration of subject matter.

TCK is the knowledge of how to leverage technology for representing and creating content through various means for a given subject area.

TPACK is the knowledge of how content areas, pedagogies, and technologies interact when teaching within a subject area and selecting suitable methods and technologies.

Although the TPACK framework has limitations (Chai et al. 2011, 2013; Kimmons 2015), it has been the most frequently applied model in educational technology research

(Ottenbreit-Leftwich and Kimmons 2018). Studies looking at TPACK have been primarily conducted in the United States, but have more recently begun to incorporate global perspectives as well (Chai et al. 2013). It has been studied in a great variety of content areas being taught: mathematics (Patahuddin et al. 2016), engineering (Jaikaran-Doe and Doe 2015), social studies (Curry and Cherner 2016), educational technology (Jaipal and Figg 2010), and science (Kramarski and Michalsky 2010). A majority of studies have been interventions (Chai et al. 2013), but methodologies have also included case studies (Özgün-Koca et al. 2011), surveys (Johnson 2012), observation (Wetzel and Marshall 2011), document analyses (Hammond and Manfra 2009), and instrument validation (Chai et al. 2013; Harris et al. 2010). In less than a decade since TPACK's introduction (Koehler and Mishra 2009), more than 600 articles have been based on the framework, and over 141 instruments devised to measure its constructs (Koehler et al. 2014). The current study's outcomes, therefore, were grounded in TPACK and measured with previously validated TPACK-based instruments.

Methodology

Context and participants

This study was conducted with 32 preservice teachers in the inclusive early childhood and elementary education programs at a mid-sized university in the Northeast. All participants were enrolled in the second of three required technology integration courses during the 2017 spring and 2017 fall semesters. The course was a one credit hour course, and participants had all successfully completed the prerequisite one credit hour course. As a one credit course, participants met 6 times during the semester for 2.5 hours per class meeting. Interspersed between these class meetings were weeks when the participants were observing and implementing lessons in elementary classrooms.

Of the 20 participants in the F2F group in the spring semester, 16 were junior-level students, three were seniors, and one was a sophomore. Seventeen of the F2F participants were female, and three were male. All 12 participants in the Flipped group, enrolled during the fall semester, were female. Eleven participants in this group were junior-level and one was a senior. Pre-tests were administered to control for potential group differences such as gender and year of school. Excluded from the study were one student who did not complete the course and one student from a different program of study.

Research design

This study implemented a 3-way mixed factorial research design with pre- and post-intervention (i.e., PrePost-I) tests to examine the impact of the FPI on preservice teachers' self-perceptions and application of TPACK in two different research conditions (i.e., Group), a flipped and a F2F course. The F2F and flipped course sections, both designed according to the operationalized FPI in Table 1, served as the interventions. Preservice teachers' TPACK development was repeatedly measured using the 7-subscale Survey of Preservice Teachers' Knowledge of Teaching and Technology (SPTKTT; Schmidt et al. 2010). This resulted in a 2 (PrePost-I) \times 2 (Group) \times 7 (Measure) mixed factorial design, with PrePost-I and Measure as within-factors and Group as the between-factor.

Table 1 Operationalization of FPI within F2F and flipped course designs

Instructional principle	F2F design	Flipped design
<p>Problem-centered: Learners engage in solving problems within an authentic context</p>	<p>Instruction was framed around the problem of integrating technology into a lesson. Learners individually designed and implemented a technology-integrated lesson as a final assessment of the whole problem. Component tasks, elements of the larger problem, supported skill and knowledge development for the problem. Tasks included the creation of digital materials and the crafting of learning objectives and assessments</p> <p>Learners were given component tasks to complete with their group as "homework." The instructor responded to questions via e-mail. The task was first shown to learners during class as the instructor modeled the specific component</p>	<p>The task was first shown to the learners online via worked examples, text-based explanations, or short videos. In class, learners were given 75 min to work on component tasks in class with their group and provided immediate feedback by the instructor</p>
<p>Activation: Learners' relevant prior experiences, knowledge, and skills are activated</p>	<p>Discussions, explorations of new material, and sharing prior knowledge through media creation facilitated the connection of new knowledge and skills with learners' previous experiences. These activities were generally allotted 30 min of class time</p>	<p>Initial activation activities were online before class with more in class to connect the online and in-class components. Online interactions were explorations of resources relevant to the upcoming class. Learners would also share prior knowledge via discussions</p>
<p>Demonstration: Instruction shows the learner what is being learned rather than just talking about what is being learned</p>	<p>The instructor provided worked examples of tasks, and presentations of content with video, audio, and text-based examples. The instructor modeled TPACK and explicitly attempted to draw out students' observations through course reflections. Approximately 30 min were used to demonstrate information with examples and non-examples. 45 min were devoted to modeling teaching behaviors with technology. This component actively involved learners in a technology-integrated lesson designed for K-5 students</p>	<p>In the online space, the instructor demonstrated information about the content and examples of the content via videos, text, and images. Most content demonstrations were presented online. Similar in both versions was that 45 min of class time was specifically devoted to modeling teaching behaviors with technology. Learners could also access procedural demonstrations online. The instructor posted these procedural recordings online as a reference before class to assist students with logging into new accounts or accessing online materials</p>

Table 1 (continued)

Instructional principle	F2F design	Flipped design
Application: Learners apply their newly acquired knowledge and skills	<p>The design tasks had the highest concentration of the application principle. Except for one, all these tasks were completed after the class meetings and required learners to apply their new knowledge and skills by designing digital resources, learning activities, and lesson plans. Interspersed throughout the class time were exercises requiring some application of new knowledge. For examples, learners practiced writing learning objectives with a partner, a Kahoot quiz was completed to check for understanding, and students were given time to edit a video using EdPuzzle</p>	<p>The application principle guided most of the in-person class time. Groups had 75 min to apply their new knowledge and skills by collaboratively designing lesson components and accompanying resources. The lessons that learners designed during class required them to apply specific TPACK understandings from the module's lessons. Learners also applied their newly acquired knowledge online before class time. They completed quizzes to practice applying their knowledge and assess their learning</p>
Integration: Supports learners' transfer of learning	<p>Integrating new knowledge, potentially TPACK, into their teaching practice was the goal of both course versions. Thus, integration was implemented in the design and implementation of a technology-integrated lesson for their field experience. Both courses assigned this to students as a means of encouraging the transfer of new knowledge into practice</p>	

The F2F and Flipped groups were separated by a semester to limit the diffusion of treatment between the groups. At this point in their program, students were grouped into a cohort. Although the students were often split into multiple sections for the technology integration courses, they spent a great deal of time together in their other courses. Conducting the study with both groups during the same semester may have encouraged discussion between groups about course differences and the sharing of resources between sections. Thus, the F2F course design was offered to the entire cohort in the spring semester and the flipped course design was provided to the entire cohort in the fall semester.

Hypotheses

Given both treatment groups' application of the FPI and Merrill's proposition that instruction based on the FPI will be equally effective in F2F and flipped learning environments, we posited to test the following three sets of hypotheses to answer our three research questions (see Sect. 1), respectively.

H1a The FPI applied to a F2F technology integration course will increase preservice teachers' self-perceptions of TPACK.

H1b The FPI applied to a F2F technology integration course will increase preservice teachers' application of TPACK.

H2a The FPI applied to a Flipped technology integration course will increase preservice teachers' self-perceptions of TPACK.

H2b The FPI applied to a Flipped technology integration course will increase preservice teachers' application of TPACK.

H3a There will be no significant differences between preservice teachers' self-perceptions of TPACK in the F2F and Flipped courses designed according to the FPI.

H3b There will be no significant differences between preservice teachers' application of TPACK in the F2F and Flipped courses designed according to the FPI.

Instruments

Survey of preservice teachers' knowledge of teaching and technology

Preservice teachers completed the Survey of Preservice Teachers' Knowledge of Teaching and Technology (SPTKTT) in the first and last week of the course (Schmidt et al. 2009). The SPTKTT was previously validated by the instrument's developers with internal consistency reliability for the seven TPACK subscales ranging from 0.75 to 0.92 as indicated by coefficient alpha (Schmidt et al. 2009). Recently, scholars have developed instruments based on the latent TPACK variables, and confirmatory factor analyses performed in these studies also supported the measurement-structure (Bostancıoğlu and Handley 2018; Kiray 2016). A range of methods have been applied to examine various TPACK instruments and highlight the challenges and strengths of applying this model (Chai et al. 2016; Valtonen

et al. 2017). Thus, the SPTKTT continues to inform the field through its application or adaptation (Hall 2018; Wang et al. 2018).

The SPTKTT consisted of forty-six items split into seven subscales. These subscales corresponded to the seven domains of TPACK. Items were all Likert scale items ranging from 1.0 to 5.0 with 1.0 representing strongly disagree and 5.0 representing strongly agree. All items were written in the affirmative, therefore, a strongly agree response indicated the participant's perception of having the highest level of knowledge in that area. Since the subscales varied in their length, the maximum possible summed score for each subscale also differed. The maximum possible subscale totals were the following: TK (30), CK (60), PK (35), PCK (20), TCK (20), TPK (45), and TPACK (20). There was a 100% response rate for both groups on the both administrations of the survey (F2F $n = 20$; Flipped $n = 12$).

Technology Integration Assessment Rubric

For the application of TPACK, two researchers rated all participants' Pre- and Post-Intervention (I) lesson designs using the TPACK-based Technology Integration Assessment Rubric (TIAR; Harris et al. 2010). The TIAR is composed of four components that are rated on a scale of increasing proficiency from 1 to 4: Curriculum Goals and Technologies, Instructional Strategies and Technologies, Technology Selection(s), and "Fit". Consequently, four constituted a minimum score on the TIAR, and sixteen was a maximum score. When originally validated, this instrument had an internal consistency, measured using Cronbach's alpha, of 0.911 (Harris et al. 2010) and was reviewed by TPACK experts for construct and face validities.

For scoring lessons designs in this study, the primary researcher downloaded all lessons from the learning management system, anonymized the documents, and assigned random numbers to both Pre- and Post-I lesson designs. While the primary researcher was aware of which lessons were Pre- and Post-I lesson designs due to his role as course instructor, the second rater was blind to this distinction. The lesson designs, identified by the randomly assigned numbers, were ordered from least to greatest. Researchers scored the lessons in this order by completing a digital TIAR on the Qualtrics survey platform and evaluated inter-rater reliability after completing sets of ten TIAR scores. Researchers met weekly to discuss scoring, and lessons with total scores between raters differing by greater than two were scored together during these meetings. Researchers reached consensus on these lessons and drafted memos to document the discussion. These memos were then used to better articulate the researchers' interpretations of the TIAR and to improve their future inter-rater reliability. The overall inter-rater reliability of the two researchers TIAR scores on the technology-integrated lesson plans was calculated and deemed substantial (Cohen's $\kappa = 0.716$; McHugh 2012).

Handling potentially influential observations

The identification of outliers in the SPTKTT data and a single outlier in the TIAR data prompted further analysis to deduce whether the outliers constituted influential observation (Penn State University 2018). Tate defines an *influential observation* "as an observation with excessive influence on any important results. The qualifier 'excessive' means an influence which qualitatively changes some study conclusions (1998, p. 50)." An excessive influence in this case could be due to both the extreme value of the data point and its leverage on the overall data. After calculating Cook's distance (Cook 1977) for each outlier in

Table 2 Descriptive statistics for the TIAR

Score	Group	M	SD	Min	Max	Skew	Kurtosis
Pre-I	F2F	7.78	2.77	4.0	13.0	0.42	-0.70
	Flipped	8.13	3.20	4.0	14.5	0.31	-0.25
	Flipped ^a	7.55	2.61	4.0	11.0	-0.31	-1.74
Post-I	F2F	11.88	2.46	8.0	16.0	0.09	-0.99
	Flipped	12.63	2.85	6.5	16.0	-1.01	0.46
	Flipped ^a	12.96	2.73	6.5	16.0	-1.43	2.27
Gain	F2F	4.10	3.49	-1.5	11.0	0.19	-0.73
	Flipped	4.50	4.10	-5.5	10.0	-1.25	2.35
	Flipped ^a	5.41	2.75	1.0	10.0	-.004	-.883

^aOutlier has been removed from this group

the SPTKTT gain scores, none were identified as influential observations. For the TIAR gains scores, however, one observation clearly deviated from the others, with the Cook's distance value of 0.46 ($> .40$) indicating a potentially influential observation. Based on recommendations from the literature (Tate 1998), we conducted the sensitivity analyses of TIAR data with and without this influential observation (see Tables 2 and 5).

Results

Descriptive statistics

The Appendix Table 6 displays the Cronbach's alpha values for each subscale and the inter-item correlation statistics. Of the 22 Pre- and Post-I alpha values calculated for the subscales, 21 indicated a strong degree of internal consistency among the items within each scale ($\alpha > 0.80$). The Post-I TPACK construct ($\alpha = 0.782$) was the only alpha value below 0.80. Although lower, researchers have noted 0.70 as an acceptable value and have reported comparable alpha values for the SPTKTT with similar participants and contexts (L. Johnson 2012; Schmidt et al. 2009). Additionally, a 7×7 correlation/variance-covariance table included in Appendix Table 7 offers further information about the relationship of the TPACK's seven subscales and may be useful for confirmatory factor analysis and meta-analytical synthesis research in the future. Figure 1 displays the mean values for the Pre- and Post-I SPTKTT responses for each group. Additional statistics, such as the standard deviations, skew, and kurtosis were calculated from the responses for each subscale score, and can be found in the Appendix Table 8. Mean values for both groups were higher following the intervention for all TPACK domains.

From the SPTKTT's descriptive statistics, it is evident that both groups of participants had comparatively high perceptions of their TPACK at the beginning of the intervention. This can be seen more clearly in Fig. 1 as each subscales' bar on the Pre-I administration was above the midpoint of 2.5. Other than TK and TCK, all variables' means on the Pre-I administration were above 3.5. Some preservice teachers in both groups scored the maximum possible on PK, TCK, TPK, and TPACK during the Pre-I administration, and only three preservice teachers responded to items with the minimum score of one. Scoring the maximum value on the Pre-I survey eliminates the measurement of potential growth and may have limited the results of this study. Nevertheless, participants in both groups

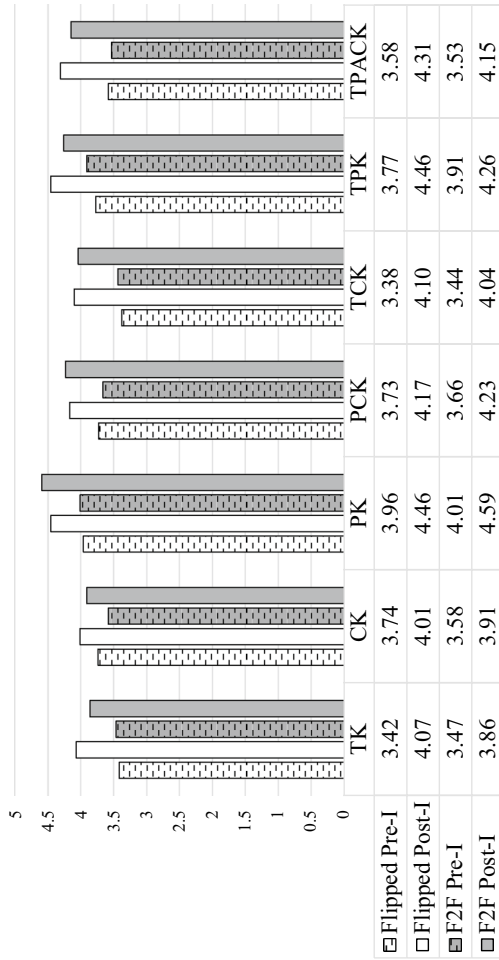


Fig. 1 Mean comparisons of pre- and post-self-perceptions of TPACK by group

Table 3 F2F and Flipped-TPACK 2×2×7 Mixed MANOVA Results

Source		Value	<i>F</i>	Sig	Partial eta squared	Observed power
PrePost-I	Pillai's Trace	.503	30.311	< .001***	.503	1.000
	Wilk's Lambda	.497	30.311	< .001***	.503	1.000
TPACK Measures	Pillai's Trace	.630	7.105	< .001***	.630	.997
	Wilk's Lambda	.370	7.105	< .001***	.630	.997
PrePost-I * Group	Pillai's Trace	.006	0.177	.677	.006	.069
	Wilk's Lambda	.994	0.177	.677	.006	.069
TPACK Measures * Group	Pillai's Trace	.078	0.353	.901	.078	.129
	Wilk's Lambda	.922	0.353	.901	.078	.129
PrePost-I * TPACK Measures	Pillai's Trace	.335	2.099	.089†	.335	.639
	Wilk's Lambda	.665	2.099	.089†	.335	.639
PrePost-I * TPACK Measures * Group	Pillai's Trace	.179	0.908	.505	.179	.290
	Wilk's Lambda	.821	0.908	.505	.179	.290

**p* < .05

***p* < .01

****p* < .001

†*p* < .10

Table 4 Single-sample T-test of gain scores for the F2F group

Variable	<i>SD</i>	M/Md	<i>t/Z</i>	<i>p</i> (1-tailed)	<i>d/r</i>
TK	5.10	2.35	2.06	.027*	0.46
PK	4.12	4.00	4.35	< .001***	0.97
CK	7.01	4.05	2.59	.009**	0.58
TCK	2.93	2.00 ^a	3.18 ^a	< .001***	0.71 ^a
PCK	2.54	6.48	11.39	< .001***	2.56
TPK	5.07	3.15	2.78	.006**	0.62
TPACK	2.42	2.00 ^b	3.18 ^b	< .001***	0.71 ^b
TIAR	3.49	4.10	5.25	< .001***	1.17

Due to a violation of normality for ^a(*SW* = .770; *p* < .001) and ^b(*SW* = .884; *p* < .021), a single sample Wilcoxon Signed Ranks Test analyzed the gain scores for TCK and TPACK. For these variables, the median (Md), Z-score, and *r*-effect size are reported. Besides Cohen's (1988) cutoffs for *d*: .20 (small), .50 (medium), .80 (large), Rosenthal (1996) added *d* = 1.30 or *r* = .70 to indicate very-large effect size

**p* < .05

***p* < .01

****p* < .001

Table 5 Single-sample T-test of gain scores for flipped group

Variable	<i>SD</i>	Mean	<i>t</i>	<i>p</i> (1-tailed)	Cohen's <i>d</i>
TK	3.75	3.92	3.62	.002**	1.05
PK	6.40	3.50	1.89	.043*	0.55
CK	5.91	3.17	1.86	.045*	0.54
TCK	3.90	2.92	2.59	.013*	0.75
PCK	4.30	5.92	4.76	.001***	1.38
TPK	8.88	6.25	2.44	.017*	0.70
TPACK	3.68	2.92	2.75	.010**	0.79
TIAR	4.10	4.50	3.81	.001***	1.10
TIAR ^a	2.75	5.41	6.53	<.001***	1.97

Besides Cohen's (1988) cutoffs for *d*: .20 (small), .50 (medium), .80 (large), Rosenthal (1996) used *d*=1.30 indicate very-large effect size

^aOutlier has been removed from this group

**p* < .05

***p* < .01

****p* < .001

visibly increased their self-perceptions of TPACK. Although the bars in this chart show preservice teachers' high self-perception of TPACK domains on the Pre-I survey, they also demonstrate consistent growth across all TPACK domains. The MANOVA with post hoc statistical tests (Tables 3, 4 and 5) will further examine the significance of this growth in Sects. 4.2 and 4.3.

For the application of TPACK, total TIAR scores (Harris et al. 2010) were calculated by combining the scores from each subscale. There were four subscales, each with a score ranging from 1 to 4. TIAR total scores, therefore, ranged from 4 -16. Table 2 displays descriptive statistics for the raters' combined TIAR scores for all participants. This combined score was obtained by calculating the mean of the two raters' scores. Gain scores were computed by calculating the difference between the Pre- and Post-I scores. This difference was then analyzed to produce the descriptive statistics in the last row for each group. The combined mean score from these descriptive statistics is used in the analyses to follow.

Differences in TPACK-based learning outcomes

To determine potential differences in TPACK-based learning outcomes and to test whether the outcomes significantly varied by group, SPTKTT scores were analyzed with a multivariate analysis of variance (MANOVA). Based on Merrill's premise that the FPI would promote learning in all contexts, we hypothesized there would be significant gains by both the F2F and Flipped groups on their self-perceptions and applications of TPACK yet no significant differences between the groups. SPSS Version 25 was used to analyze the multivariate data collected from the 3-way mixed design. The 2 × 2 × 7 design has one between-subjects factor (Group) and two within-subjects factors (Pre- and Post-Intervention and TPACK). There were seven levels in the factor of TPACK: TK, CK, PK, TCK, TPK, PCK, TPACK. Group had two levels, Flipped and F2F. Group was entered as F2F followed by Flipped. No outliers were deemed statistically

influential. Regarding the assumption of normality, the skewness values were satisfactory, but three kurtosis values were problematic with absolute values greater than two (Flipped_TK_Pre-I=2.542, Flipped_CK_Post-I=2.612, and Flipped_TPK_Post-I= -2.185). The analyses proceeded since not all variables have to be normally distributed to perform a MANOVA; slight variations of skew and kurtosis do not significantly impact the power and level of significance (Rao 2016). Pillai's Trace, however, was reported in addition to the commonly reported Wilk's Lambda in Table 3 as it is more robust to violations of such assumptions. Evaluations of the remaining assumptions of homogeneity of variances and covariances, linearity, and multicollinearity were deemed suitable to perform the MANOVA.

Based on Pillai's Trace, the multivariate effect was statistically significant by the Pre-Post-I outcome variables, $F(1, 30) = 30.311, p < 0.001, \eta^2 = 0.503$. The multivariate test further indicated there were no significant interactions between Group and PrePost-I. Thus, the PrePost-I result demonstrates that participants in both the F2F and Flipped groups increased their TPACK self-perceptions following the intervention. The lack of an interaction effect with Group evidences that the observed increases in TPACK were not different by treatment group. The significant gains of learners in both groups lend support to the FPI's potential to design effective instruction. Furthermore, the significant gains contrasted by the absence of a statistically significant interaction between PrePost-I measures and Group fails to nullify Merrill's premise that the FPI promote learning in all contexts. These results did not discern differences in the outcomes between the two learning contexts designed by the FPI.

For the F2F and Flipped Groups (see Table 3), with Pillai's Trace, the TPACK Measures factor was statistically significant, $F(1, 30) = 7.105, p < 0.001, \eta^2 = 0.630$. This indicates that each of the seven factors within TPACK Measures were distinct. Regardless of which measure of TPACK is observed, the intervention significantly improved the post-test so that it is different than the pre. The interaction of TPACK Measures by Group was not statistically significant, thus demonstrating no interaction between TPACK Measures and respondents' grouping. Therefore, the observed significant increases across the TPACK measures was not different for the F2F and Flipped groups.

The marginally significant interaction of Measures by PrePost-I $F(1, 30) = 2.099, p < 0.089, \eta^2 = 0.335$ may be explained by possible patterns in how participants responded to measures before and after the course. As previously discussed, the main effect was significantly different before and after the intervention, and the statistical significance of each TPACK Measure was also unique. This slight interaction of PrePost-I with Measures indicates that the significant improvement following the intervention was not uniformly observed across the different TPACK Measures (see Table 4). While the gains for each TPACK Measure were significant, the degree of significance varied. Hence, the next section details the univariate posthoc analyses used to further examine individual constructs of the TPACK Measure.

FPI in the F2F and flipped courses

To better comprehend the statistically significant outcomes indicated by the MANOVA, univariate results were examined for each group. Since the assumption of normality was not met for all outcome variables, the posthoc analysis was conducted with *t*-tests for normally distributed outcome variables and a Wilcoxon Signed Ranks Test for variables violating this assumption. The Shapiro–Wilk (SW) test indicated that the assumption of normality was

violated for subscales TCK ($SW=0.770$; $p<0.001$) and TPACK ($SW=0.884$; $p=0.021$). TCK and TPACK were thus analyzed with a Wilcoxon Signed Ranks test. The results for these two variables are reported in Table 4 with the median, Z-score, and r -effect size.

The results in Table 4 indicate that participants in the F2F group significantly increased their self-perceptions TPACK in all domains, and significantly increased their application of TPACK to technology-integrated lesson designs (TIAR). This statistically significant growth, with medium to large effect sizes, offers support for the FPI's capacity to guide the design of F2F learning environments.

The MANOVA also showed self-perceptions of TPACK to be positively associated with participants in the flipped course. A single-sample t -test of gain scores offered additional insight for individual outcome variables. Table 5 displays the results for this analysis, and provides TIAR results of the sensitivity analysis with and without the influential observation previously discussed (see Sect. 3.5). Increases in self-perceptions of each TPACK domain were statistically significant with medium to large effect sizes. Preservice teachers' increased application of TPACK to technology-integrated lesson designs was also statistically significant with a large effect size. Therefore, the results supported the hypothesis that the FPI will promote development of all TPACK domains in both flipped and F2F courses.

Discussion

Effective design is effective design

In both the F2F and Flipped groups, preservice teachers' self-perceptions and application of TPACK statistically significantly increased. The strength of their growth in application of TPACK to technology-integrated lesson designs (F2F $p<0.001$, $d=1.17$; Flipped^a $p<0.001$, $d=1.97$) and the consistency of their growth in self-perceptions of all TPACK Measures $F(1, 30)=30.311$, $p<0.001$, $\eta^2=0.503$ provides compelling support for the FPI's potential to guide effective course design. The absence of statistically significant differences between the F2F and Flipped groups on any measure suggest that the FPI may have been no more or less effective when applied to designing flipped and F2F courses.

Excerpts from participants' course reflections offer insight into how they perceived the FPI-based course experiences. Describing interactions with the FPI's problem-centered principle and its problem progression corollary as well as the application and integration principles, a participant in the Flipped group wrote the following in a course reflection:

The way [the course] is organized allows for scaffolded guidance. For example, for the first lesson, we were provided both the learning objectives and the lesson standards. This acted as the introduction to the technological concepts, and we were given the most support. Looking at this last lesson, our design teams wrote our own learning objectives and chose our own lesson standards. This scaffolded approach, beginning with the most support and ending with the least support, helped us learn how to plan lessons with integrated technology.

In retelling experiences with the "scaffolded guidance" of this problem-centered design, the participant contrasted the final design experience with the first design experience. The participant explained what was provided in the first lesson and how they could still succeed amidst the waning support. Another participant noted in a reflection the benefits of

the application and demonstration phases of the course design. “It is also helpful to try out these different means of educational technology and apply them into a lesson so we have more experience with actually writing a lesson plan that utilizes these technologies. Practicing them in class also helps us become more confident in using it in a classroom.” Designing opportunities for application in class and gradually releasing responsibility to students amidst an increasingly complex problem was perceived as beneficial to their learning experience.

Thus, these results align with Merrill’s premise that the FPI have potential to inform effective instruction in varied disciplines, contexts, and learning environments (Merrill 2012). While technologies, environments, and educational terminology may vary, he posits that there remain fundamental instructional principles and strategies. “While their implementation may be radically different, those learning strategies that best promoted learning in the past are those learning strategies that will best promote learning in the future (Merrill 2012).” This study substantiates his claim regarding the effectiveness of the FPI and their potential for designing F2F and blended approaches. These results may therefore extend support for the importance of strategies for learning and instruction amidst the vicissitudes of technology adoptions and emerging environments (Lo 2018).

Potential influence of design variations

Although not statistically significantly different, there were noteworthy differences between the F2F and Flipped groups’ effect sizes of select variables. For example, the Flipped group’s TK effect size was more than double the magnitude of the F2F group (Flipped $d=1.05$; F2F $d=0.46$). Both effect sizes indicate a moderate to critical practical importance of the intervention yet the variance in magnitude elicits further consideration. From a design perspective, the differences in TK growth may be explained by how the FPI’s activation principle was applied uniquely to the F2F and flipped courses.

Before discussing the affordances of each design, we will present a brief example of an instructional segment based on the activation principle’s corollary of providing learners with opportunities to gain relevant new experiences as a foundation for learning new knowledge and skills. As a means of gaining new experiences with Web 2.0 tools during a module focused on integrating Web 2.0 tools into instruction, learners in the Flipped group completed the following online activities before class.

1. Read overview of Web 2.0 tools, module objectives, and directions
2. Watched two video segments (<5 min each). One focused on the differences between web generations and the other was a fifth-grade teacher sharing five ways in which Web 2.0 tools were integrated into instruction
3. Explored a mashup of Web 2.0 tools
4. Read an assigned case about a teacher’s experience integrating Web 2.0 tools
5. Viewed an instructor-created VoiceThread that discussed instructional strategies for effectively integrating technology
6. Via a web-based class bulletin board, responded to a discussion prompt about their Web 2.0 explorations and shared ideas for future application.
7. Completed a quiz that assessed the module’s learning objectives

Integrated into a presentation at the beginning of class, learners in the F2F group read the same overview of Web 2.0 tools and viewed the same video in class that the Flipped group was assigned online. The F2F group was also given time to explore the mashup of Web 2.0 tools, discuss their explorations within small groups and share highlights of these explorations with the whole class. Although provided with similar content and opportunities to gain new experiences in both designs, learners' outcomes were potentially influenced by the varying interactions with content rendered by the F2F and flipped approaches. Describing the flipped section in a reflection, Bridget wrote, "I found the activities that I completed online before coming to class were very helpful. Due to the fact that I had read about Google Sites prior to coming to class, I was able to understand and incorporate the use of Google Sites into the lesson plan." While both course designs included activation through gaining experience with new knowledge, the flipped participants, by design, encountered these experiences primarily online before class.

In the F2F group, learners gained new experience with digital tools during the initial part of the class meeting. Allotted time and afforded autonomy, therefore, were uncontrolled variables in each design. For example, 20 min was the maximum amount of time afforded to the activation phase instructional activities with the F2F group. F2F participants wrote of being introduced to the technologies in class. For example, when asked about when he explored new technologies introduced by the course, Andrew said, "Probably in class. Yeah I mean we were really busy outside of class. I feel like we had a lot of time to do it more in class." Learners in the Flipped group, however, managed their own time when engaged in these activation phase activities as they completed the assignments online.

Gaining new experiences prior to the class meeting provided a preview of the learning objectives for the Flipped group and may have more effectively activated learners' relevant schema. A participant in the Flipped group wrote, "I am always so grateful that we use the modules outside of class as a way to familiarize ourselves with content prior to class," and another participant commented that the flipped design "was comforting in a way cause I knew what to expect. I knew it was happening. But its also because I knew why I was doing something." Learners arrived at class with foundational knowledge, relevant new experiences, and an awareness of what content the instructor would accentuate during class time.

As for autonomy, the Flipped group controlled the pace, place, and time when they would complete and submit assignments prior to the class meeting while the F2F group completed and submitted the work in class. Autonomy, as compared to the Flipped group's activation phase, may have been constrained for the F2F group in terms of pacing, group size, and location. Participants in the Flipped group were freer to choose how these interactions occurred. They could select when to interact, where to interact, the duration and frequency of the interactions, or could refrain from interacting at all. The choice not to engage new tools was open to learners who were already familiar with the digital tools and could satisfactorily demonstrate their mastery via the assessment. The Flipped group learners could also select to gain experience with the new digital tools individually, with a partner, as a group, or even as a class. Their constraints were minimal when compared to the F2F group who were assigned time, place, and directions for grouping.

Grounding their work in self-determination theory, Abeysekera and Dawson (2015) posit that the flipped approach may increase students' motivation by satisfying their needs for autonomy and competence. It may be that learners in this study's Flipped group sensed greater autonomy and competence and were more motivated to learn about

technology. As the previous paragraphs detailed, many facets of the FPI's implementation in the flipped approach may have better supported these factors. The uncontrolled variations of time on task and afforded autonomy resulting from the design distinctions may have contributed to the noticeable differences in participants' TK growth indicated by the much larger effect size for the Flipped group.

Implications for future research and course design

The field of instructional design and technology should carefully attend to how instructional interventions and their comparisons are designed and reported. In the excitement for innovative approaches, it can seem effective to refer to studies as comparisons of *emerging* with *traditional* when we are actually studying novel variations of foundational instruction with approaches long viewed as ineffective (Callison 2015; Touchton 2015). This study's non-significant differences between groups are consistent with Lo et al.'s (2018) comparison of four FPI-based flipped courses. In their study, historical course data was used as the comparison group. Of the four classes, one exhibited similar outcomes when comparing the FPI-based flipped course and its historical F2F counterpart. Unique to this course, the teacher embedded active learning and interactive components in the F2F lecture. The teacher commented that little changed when flipping the course with the FPI (Lo et al. 2018). Hence it is difficult to conclude whether learning gains observed with the other groups, whose instruction prior to the flip was primarily direct lecture, can be attributed to the FPI, a flipped approach, replacing the lecture with active learning, or some combination.

Similar incompatibilities within research on flipped classrooms has led some to claim the flipped approach as superior to F2F instruction (Kurt 2017). Albeit a single study, in one discipline, and a specific context, this research examined differences between two rigorously designed courses that both applied the same principles of instruction. The lack of significant differences between groups is not something to be disregarded as results with no bearing. They represent the potential of the FPI to inform instructional design in multiple contexts, and they are a call to more critically design comparisons of flipped approaches.

As the use of the FPI in flipped instruction examined by this study was intended for design beyond the teacher preparation context, it would be beneficial for researchers in other disciplines to study the efficacy of designing with it and the impact on learning outcomes. Does the premise of the FPI's applicability to diverse contexts and emerging environments hold true? There have been ineffective implementations (Cargile and Karkness 2015), inadequately conceptualized designs (O'Flaherty and Phillips 2015) and struggles with designing flipped courses reported throughout the literature. Following the FPI when designing a flipped course could be a valuable approach for skilled designers and novices alike. It affords a flexible approach to design, provides supportive prescriptions, and offers a much-needed conceptual framework for bridging pre- and in-class activities.

Contributions to theoretical understandings

Findings from this study fail to disconfirm Merrill's claims regarding the FPI. "First, learning from a given program will be promoted in direct proportion to its implementation of first principles. Second, first principles of instruction can be implemented in any delivery system or using any instructional architecture (Merrill 2002, p. 44)." While this

study did not assign a value to the degree of the FPI's implementation in each course, the intent was to equivalently apply the FPI in both F2F and flipped approaches. The documentation of course designs and manipulation checks attempted to facilitate equivalent implementation. The non-significant differences between the F2F and Flipped group's quantitative outcomes, therefore, is consistent with Merrill's proposition. It is also noteworthy that the participants' increases were larger in this study than a pilot study (Hall 2018). While several confounding factors warrant caution when comparing these studies, the intent of the current study was to increase the strength of the intervention by implementing the FPI with greater fidelity than was done in the pilot study. Although not providing confirmatory evidence, had the current study's course outcomes been drastically different from each other or indicated lesser gains or even equivalent results to the prior iteration, it would be cause for questioning the proposed property of proportionate implementation and learning.

The second property, denoting FPI's universal applicability, claims the principles apply to any instructional environment (Merrill 2012). This study and others have tested this claim and in flipped courses (Hall 2018; Lo and Hew 2017; Lo et al. 2018). Clearly, the presence of studies testing the FPI in various environments evidence that it can be applied to these environments. Yet, knowing something can be done does not indicate that it should be done and there is yet more research needed on the impact on learning outcomes in FPI designed courses. Similarly, this study's positive findings of participants' TPACK growth does not exclude the possibility of stronger results occurring had another instructional design model been applied.

Given the broad nature of the principles and their corollaries, Merrill notes they can be easily misinterpreted and applied (Merrill 2012). One may also argue, as Merrill did, that the FPI are well supported by what is already known about learning and instruction. Rather than focusing on the general principles, which may lead to vague designs supporting widely applicable and commonly accepted rules, it may be better for future studies to focus on specific prescriptions for implementing the principles. Are there certain prescriptions that more effectively leverage the FPI? What FPI-based prescriptions best support learning in specific environments? How do learners experience these varying prescriptions? Emphasizing prescriptions in future studies will likely shift attention from the FPI's universality to situationalities (Reigeluth 2013), and this alteration may result in more finely tuned, context specific, and replicable studies that extend our knowledge of instructional design.

Limitations

Unmeasured confounding variables presented potential limitations in this study and the comparability of the two groups. An assumption of the course design was an approximately even amount of time would be needed to complete the assignments outside of class in both treatment groups. The design intended for learners to spend 45 – 60 min working on the pre-class activities (Flipped) or their design team tasks outside of class (F2F). Time spent on these course activities, however, was not monitored and could have differed between interventions (Means et al. 2013).

Another limitation of the coursework outside of class was the expectation that F2F students would complete their design team tasks as a team. A tenet of constructionism and a corollary of the FPI's application principle is to collaborate with peers, thus constituting an important design element. Both groups collaborated with peers during the model lessons in

class. The Flipped group collaborated during the design team or problem progression tasks in class as well. As these tasks were assigned to the F2F group to complete together for homework, they were more open to negotiating this collaborative arrangement. The degree to which they collaborated similar to the Flipped group represents another unaccounted factor. Future studies could assign individual tasks for homework or collect data that would inform how groups worked together outside of class.

In future studies, a contrasting control group designed by an instructional design model entirely unique from the FPI would facilitate more robust results. This presents a theoretical dilemma, however, as Merrill posits that the FPI represent principles from a large sample of instructional design models (2012). If the FPI are truly universal to all good instructional models as Merrill claimed, a control condition exclusive of any FPI elements may be purposefully designing ineffective instruction and represent a potentially harmful approach to research. This dilemma may be worth solving, however, as it would provide a more valid representation of differences between groups.

Conclusion

The significance of this study was the examination of the FPI applied to both a F2F and flipped course, and the use of a flipped course to develop preservice teachers' TPACK development. In the Flipped group, the statistically significant growth of preservice teachers' TPACK self-perceptions and application inform a literature gap that had been previously identified as lacking empirical support of robust educational outcomes (Kurt 2017; O'Flaherty and Phillips 2015). Comparing a flipped course with a F2F course designed according to the same instructional principles also uniquely contributes to what is known about the efficacy of flipped courses. The significant relationship of the FPI with positive TPACK-based learning outcomes in both groups suggest strong potential for informing future preservice teacher technology integration development.

In a world that is constantly changing, this study lent support to instructional design principles built upon the premise that effective, efficient, and engaging instruction may not be radically different today than it was before the term "flipped" was ever conceived. While this study sought to develop preservice teachers' self-perceptions and application of TPACK, it seems to also have effectively modeled educational technology well. Educational technology has been defined as "the disciplined application of scientific principles and theoretical knowledge to support and enhance human learning and performance (Spector et al. 2008, p. 820). Since educational technology as an application of principles and theoretical knowledge was integrated in both course designs to support participants' TPACK development, it may not be surprising that both groups' outcomes were significant yet not significantly different.

Acknowledgements We are deeply grateful to Dr. Jiaming Cheng who supported this study by observing the F2F and Flipped course implementations and by assisting with the analysis of the technology integrated lesson plans. We also thankfully acknowledge the funding provided by the Syracuse University School of Education Research and Creative Grant Program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

See Tables 6, 7, and 8.

Table 6 Reliability statistics for SPTKTT and TIAR in this study

Instrument	Subscales	Cronbach's α		Inter-item correlation		Average inter-item correlation		# of items
		Pre-I	Post-I	Pre-I	Post-I	Pre-I	Post-I	
SPTKTT	TK	0.897	0.916	.43-.75	.50-.74	0.60	0.66	6
	CK Math	0.918	0.940	.78-.83	.83-.85	0.81	0.85	3
	CK SS	0.923	0.838	.77-.88	.53-.70	0.82	0.63	3
	CK Science	0.933	0.933	.76-.86	.80-.87	0.82	0.82	3
	CK Literacy	0.943	0.963	.81-.89	.85-.95	0.85	0.90	3
	PK	0.924	0.911	.31-.85	.39-.93	0.65	0.62	7
	PCK	0.919	0.836	.65-.82	.35-.82	0.75	0.59	4
	TCK	0.977	0.909	.87-.93	.58-.93	0.91	0.71	4
	TPK	0.956	0.939	.52-.93	.34-.85	0.72	0.64	9
	TPACK	0.953	0.782 ^a	.70-.93	.25-.74	0.81	0.52	4
TIAR	N/A	0.903	0.905	.59-.85	.61-.80	0.72	0.71	4

^a*F*-test (McGraw and Wong 1996) showed that this reliability is not statistically different from the most commonly used cutoff value of .80 ($F = .917$, $df_1 = 31$, $df_2 = 93$, $p = .596$)

Table 7 Pre- (post-) intervention subscale correlations and variance-covariances

Subscales	TK	CK	PK	PCK_	TCK	TPK	TPACK
TK							
Pre-I (Post-I) <i>r</i> ^a	1 (1)						
Pre-I (Post-I) <i>cov</i> ^b	.555 (.500)						
CK							
Pre-I (Post-I) <i>r</i>	.335 (.519**)	1 (1)					
Pre-I (Post-I) <i>cov</i>	.146 (.190)	.344 (.269)					
PK							
Pre-I (Post-I) <i>r</i>	.402* (.371*)	.253 (.285)	1 (1)				
Pre-I (Post-I) <i>cov</i>	.167 (.122)	.083 (.069)	.313 (.218)				
PCK							
Pre-I (Post-I) <i>r</i>	.565** (.381*)	.504** (.525**)	.671** (.498**)	1 (1)			
Pre-I (Post-I) <i>cov</i>	.304 (.148)	.214 (.149)	.272 (.128)	.524 (.300)			
TCK							
Pre-I (Post-I) <i>r</i>	.433* (.497**)	.389* (.657**)	.311 (.433*)	.685** (.550**)	1 (1)		
Pre-I (Post-I) <i>cov</i>	.296 (.206)	.210 (.200)	.160 (.118)	.456 (.176)	.845 (.343)		
TPK							
Pre-I (Post-I) <i>r</i>	.697** (.552**)	.521** (.389**)	.735** (.494**)	.772** (.269)	.555** (.647**)	1 (1)	
Pre-I (Post-I) <i>cov</i>	.327 (.202)	.192 (.104)	.259 (.119)	.352 (.076)	.322 (.196)	.397 (.268)	
TPA-CK							
Pre-I (Post-I) <i>r</i>	.466** (.510**)	.553** (.618**)	.354* (.480**)	.691** (.569**)	.753** (.603**)	.698** (.612**)	1 (1)
Pre-I (Post-I) <i>cov</i>	.265 (.178)	.247 (.158)	.151 (.110)	.382 (.153)	.528 (.174)	.336 (.156)	.582 (.242)

^aPre-I (Post-I) *r*: pre-intervention (post-intervention) Pearson Correlation *r*

^bPre-I (Post-I) *cov*: pre-intervention (post-intervention) covariance. **p* < 0.05. ***p* < 0.01. *N* = 32

Table 8 Descriptive statistics for SPTKTT

Variable	Time	Group	M	SD	Raw scores				
					Md ^a	Min	Max	Skew	Kurtosis
TK	Pre-I	F2F	3.47	0.63	21.0	13.0	28.0	-0.059	-0.232
		Flipped	3.42	0.94	21.0	7.0	30.0	-0.944	2.542
	Post-I	F2F	3.86	0.75	23.5	13.0	30.0	-0.337	0.225
		Flipped	4.07	0.64	24.0	17.0	30.0	-0.396	0.197
CK	Pre-I	F2F	3.58	0.64	43.0	32.0	60.0	0.399	-0.470
		Flipped	3.74	0.5	45.5	36.0	56.0	-0.005	-0.339
	Post-I	F2F	3.91	0.58	48.0	31.0	60.0	-0.361	0.606
		Flipped	4.01	0.41	48.0	42.0	60.0	1.384	2.612
PK	Pre-I	F2F	4.01	0.53	28.0	23.0	35.0	0.606	-0.234
		Flipped	3.96	0.63	27.5	21.0	35.0	0.079	-0.363
	Post-I	F2F	4.59	0.47	33.5	25.0	35.0	-0.977	-0.438
		Flipped	4.46	0.48	32.0	26.0	35.0	-0.138	-1.594
PCK	Pre-I	F2F	3.66	0.73	14.0	10.0	20.0	0.609	-0.268
		Flipped	3.73	0.74	16.0	9.0	20.0	-0.404	0.328
	Post-I	F2F	4.23	0.62	20.0	15.0	20	-0.316	-0.567
		Flipped	4.17	0.43	20.0	18.8	20	1.388	0.735
TCK	Pre-I	F2F	3.44	0.93	14.0	4.0	20.0	-0.708	1.473
		Flipped	3.38	0.93	12.0	8.0	20.0	0.738	-0.170
	Post-I	F2F	4.04	0.67	16.0	12.0	20.0	-0.048	-0.782
		Flipped	4.10	0.45	16.0	15.0	20.0	1.553	1.380
TPK	Pre-I	F2F	3.91	0.55	36.0	27.0	45.0	0.225	0.204
		Flipped	3.77	0.76	35.0	24.0	45.0	0.295	-0.650
	Post-I	F2F	4.26	0.54	38.0	27.0	45.0	-0.359	-0.036
		Flipped	4.46	0.48	40.0	35.0	45.0	0.028	-2.185
TPACK	Pre-I	F2F	3.53	0.78	14.5	8.0	20.0	0.095	-0.040
		Flipped	3.58	0.77	15.5	8.0	20.0	-0.331	0.789
	Post-I	F2F	4.15	0.48	16.0	12.0	20.0	0.134	1.258
		Flipped	4.31	0.52	16.0	15.0	20.0	0.601	-1.640

^aMedian

References

- Abeyssekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14.
- After Moore's Law: The future of computing. (2016). *The Economist*. Retrieved from <https://www.economist.com/news/leaders/21694528-era-predictable-improvement-computer-hardware-ending-what-comes-next-future>
- Andre, T. (1997). Selected microinstructional methods to facilitate knowledge construction: Implications for instructional design. In R. D. Tennyson, F. Schott, N. Seel, & S. Dijkstra (Eds.), *Instructional-design: International perspective: Theory, research, and models* (pp. 243–267). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Angeli, C., Valanides, N., & Christodoulou, A. (2016). Theoretical considerations of technological pedagogical content knowledge. In M. C. Herring, M. J. Koehler, & P. Mishra (Eds.), *Handbook of technological pedagogical content knowledge (TPACK) for educators* (2nd ed., pp. 11–32). New York: Routledge.

- Bednar, A., Cunningham, D., Duffy, T., & Perry, J. D. (1991). Theory into practice: How do we link? In G. J. Anglin (Ed.), *Instructional technology: Past, present, and future* (pp. 88–101). Berlin: Springer.
- Bellini, J. L., & Rumrill, P. D. (1999). Perspectives on scientific inquiry: Validity in rehabilitation research. *Journal of Vocational Rehabilitation, 13*, 131–138.
- Beloit College. (2016). The Beloit College mindset list for the class of 2020. Retrieved September 19, 2016, from <https://www.beloit.edu/mindset/2020/>
- Bostancıoğlu, A., & Handley, Z. (2018). Developing and validating a questionnaire for evaluating the EFL ‘Total PACKage’: Technological pedagogical content knowledge (TPACK) for English as a foreign language (EFL). *Computer Assisted Language Learning, 31*(5–6), 572–598.
- Callison, D. (2015). Classic instructional notions applied to flipped learning for inquiry. *School Library Monthly, 31*(6), 20–22.
- Cargile, L. A., & Karkness, S. S. (2015). Flip or flop: Are math teachers using Khan Academy as envisioned by Sal Khan? *TechTrends, 59*(6), 21–27.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2011). Exploring the factor structure of the constructs of technological, pedagogical, content knowledge (TPACK). *The Asia-Pacific Education Researcher, 20*(3), 595–603.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2013). A review of technological pedagogical content knowledge. *Educational Technology & Society, 16*(2), 31–51.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2016). A review of quantitative measures of technological pedagogical content knowledge (TPACK). In M. C. Herring, M. J. Koehler, & P. Mishra (Eds.), *Handbook of Technological Pedagogical Knowledge (TPACK) for Educators* (Second, pp. 87–106). New York: Routledge.
- Cheung, W. S., & Hew, K. F. (2015). Applying “First Principles of Instruction” in a blended learning course. In K. C. Li, T. L. Wong, S. K. Cheung, J. Lam, & K. K. Ng (Eds.), *Technology in education: Transforming educational practices with technology* (pp. 127–135). Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-46158-7>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Coles, P., Cox, T., Mackey, C., & Richardson, S. (2006). *The toxic terabyte: How data-dumping threatens business efficiency*. IBM Global Technology Services.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 42*(1), 65–68.
- Creswell, J. W. (2012). *Education research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Saddle River, NJ: Pearson Education Inc.
- Curry, K., & Cherner, T. (2016). Social studies in the modern era: A case study of effective teachers’ use of literacy and technology. *The Social Studies, 107*(4), 123–136. <https://doi.org/10.1080/00377996.2016.1146650>
- Donaldson, J. A. (2017). Emerging technology: Instructional strategies for nailing Jell-O to a tree. In Y. Li, M. Zhang, C. J. Bonk, & W. Zhang (Eds.), *Learning and knowledge analytics in open Education* (pp. 89–97). New York: Springer.
- Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6*(4), 50–72.
- Frick, T. W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2009). College student perceptions of teaching and learning quality. *Educational Technology Research and Development, 57*(5), 705–720. <https://doi.org/10.1007/s11423-007-9079-9>
- Frick, T. W., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research and Development, 58*(2), 115–136. <https://doi.org/10.1007/s11423-009-9131-z>
- Gardner, H., & Davis, K. (2014). *The app generation: How today’s youth navigate identity, intimacy, and imagination in a digital world*. New Haven, CT: Yale University Press.
- Gardner, J., & Belland, B. R. (2012). A conceptual framework for organizing active learning experiences in biology instruction. *Journal of Science Education and Technology, 21*(4), 465–475.
- Hall, J. A. (2018). Flipping with the first principles of instruction: An examination of preservice teachers’ technology integration development. *Journal of Digital Learning in Teacher Education, 34*(4), 201–218. <https://doi.org/10.1080/21532974.2018.1494520>
- Hall, J. A., & Lei, J. (2020). Conceptualization and application of a model for flipped instruction: A design case within teacher education. *Research Issues in Contemporary Education, 5*(2), 24–54.
- Hammond, T. C., & Manfra, M. M. (2009). Digital history with student-created multimedia. *Social Studies Research and Practice, 4*(3), 139–150.

- Harris, J., Grandgenett, N., & Hofer, M. J. (2010). Testing a TPACK-based technology integration assessment rubric. In C. D. Maddux, D. Gibson, & B. Dodge (Eds.), *Research Highlights in Technology and Teacher Education 2010* (pp. 323–331). Society for Information Technology and Teacher Education. Retrieved from <https://publish.wm.edu/bookchapters/6>
- Hoffman, E. S. (2014). Beyond the flipped classroom: Redesigning a research methods course for e3 instruction. *Contemporary Issues in Education Research*, 7(1), 51–62.
- Jaikaran-Doe, S., & Doe, P. E. (2015). Assessing technological pedagogical content knowledge of engineering academics in an Australian regional university. *Australasian Journal of Engineering Education*, 20(2), 157–167. <https://doi.org/10.1080/22054952.2015.1133515>
- Jaipal, K., & Figg, C. (2010). Unpacking the “Total PACKage”: Emergent TPACK characteristics from a study of preservice teachers teaching with technology. *Journal of Technology and Teacher Education*, 18(3), 415–441.
- Johnson, L. (2012). *The Effect of Design Teams on Preservice Teachers' Technology Integration*. Syracuse University.
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hal, C. (2016). *NMC horizon report: 2016 higher education edition*. Austin, Texas.
- Jonassen, D. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 215–239). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Kimmons, R. (2015). Examining TPACK's theoretical future. *Journal of Technology and Teacher Education*, 23(1), 53–77.
- Kiray, S. A. (2016). Development of a TPACK self-efficacy scale for preservice science teachers. *International Journal of Research in Education and Science*, 2(2), 527–541.
- Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The technological pedagogical content knowledge framework. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., Vol. 13, pp. 260–263). New York: Springer. <https://doi.org/https://doi.org/10.1007/978-1-4614-3185-5>
- Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, 20(5), 434–447. <https://doi.org/10.1016/j.learninstruc.2009.05.003>
- Kurt, G. (2017). Implementing the flipped classroom in teacher education: Evidence from Turkey. *Educational Technology & Society*, 20(1), 211–221.
- Lee, S., & Koszalka, T. A. (2016). Course-level implementation of first principles, goal orientations, and cognitive engagement: A multilevel mediation model. *Asia Pacific Education Review*, 17(2), 365–375.
- Lo, C. K. (2018). Grounding the flipped classroom approach in the foundations of educational technology. *Educational Technology Research and Development*, 66(3), 1–19. <https://doi.org/10.1007/s11423-018-9578-x>
- Lo, C. K., & Hew, K. F. (2017). Using “First Principles of Instruction” to design secondary school mathematics flipped classroom: The findings of two exploratory studies. *Educational Technology & Society*, 20(1), 222–236.
- Lo, C. K., Lie, C. W., & Hew, K. F. (2018). Applying “First Principles of Instruction” as a design theory of the flipped classroom: Findings from a collective study of four secondary school subjects. *Computers & Education*, 118, 150–165. <https://doi.org/10.1016/j.compedu.2017.12.003>
- Margulieux, L. E., McCracken, W. M., & Catrambone, R. (2016). A taxonomy to define courses that mix face-to-face and online learning. *Educational Research Review*, 19, 104–118.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Means, B., Toyoma, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47.
- Merriënboer, J. J. G., Clark, R. E., & Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50(2), 39–61.
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59.
- Merrill, M. D. (2012). *First principles of instruction* (1st ed.). San Francisco: Pfeiffer.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.

- Nelson, K. R. (2015). Application of Merrill's first principles of instruction in a museum education context. *Journal of Museum Education*, 40(3), 304–312.
- Nelson, L. M. (1999). Collaborative problem solving. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 241–267). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, 25, 85–95.
- Ottenbreit-Leftwich, A., & Kimmons, R. (Eds.). (2018). *The K-12 educational technology handbook*. EdTech Books. Retrieved from <https://edtechbooks.org/k12handbook>
- Özgün-Koca, S. A., Meagher, M., & Edwards, M. T. (2011). A teacher's journey with a new generation handheld: Decisions, struggles, and accomplishments. *School Science and Mathematics*, 111(5), 209–224. <https://doi.org/10.1111/j.1949-8594.2011.00080.x>
- Patahuddin, S. M., Lowrie, T., & Dalgarno, B. (2016). Analysing mathematics teachers' TPACK through observation of practice. *Asia-Pacific Education Researcher*, 25(5–6), 863–872. <https://doi.org/10.1007/s40299-016-0305-2>
- Penn State University. (2018). Lesson 11: Influential points. Retrieved April 26, 2018, from <https://onlinecourses.science.psu.edu/stat501>
- Rao, C. R. (2016). Multivariate analysis of variance. In R. E. Schumacker (Ed.), *Using R with multivariate statistics: A primer* (pp. 57–80). Thousand Oaks, CA: Sage.
- Reigeluth, C. M. (2013). Instructional theory and technology for the new paradigm of education. *The F.M. Duffy Reports*, 18(4), 1–21.
- Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, 21(4), 37–59.
- Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 161–181). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Schmidt, D. A., Baran, E., Thompson, A. D., Koehler, M. J., Mishra, P., & Shin, T. S. (2010). Survey of preservice teachers' knowledge of teaching and technology, 1–8.
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK). *Journal of Research on Technology in Education*, 42(2), 123–149.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *American Educational Research Association*, 15(2), 4–14. <https://doi.org/10.1017/CBO9781107415324.004>
- Spector, J. M., Merrill, M. D., van Merriënboer, J. J. G., & Driscoll, M. P. (Eds.). (2008). *Handbook of research on educational communications and technology* (3rd ed.). New York: Lawrence Erlbaum Assoc.
- Tate, R. (1998). *An introduction to modeling outcomes in the behavioral and social sciences* (2nd ed.). Edina, MN: Burgess Publishing.
- Tiruneh, D. T., Weldeslassie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, 64(3), 481–505.
- Touchton, M. (2015). Flipping the classroom and student performance in advanced statistics: Evidence from a quasi-experiment. *Journal of Political Science Education*, 11(1), 28–44. <https://doi.org/10.1080/15512169.2014.985105>
- Valtonen, T., Sointu, E., Kukkonen, J., Kontkanen, S., Lambert, M. C., & Mäkitalo-Siegl, K. (2017). TPACK updated to measure pre-service teachers' twenty-first century skills. *Australasian Journal of Educational Technology*, 33(3), 15–31.
- Wang, W., Schmidt-Crawford, D. A., & Jin, Y. (2018). Preservice teachers' TPACK development: A review of literature. *Journal of Digital Learning in Teacher Education*, 34(4), 234–258. <https://doi.org/10.1080/21532974.2018.1498039>
- Wetzel, K., & Marshall, S. (2011). TPACK goes to sixth grade: Lessons from a middle school teacher in a high-technology-access classroom. *Journal of Digital Learning in Teacher Education*, 28(2), 73–82.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Jacob A. Hall is an Assistant Professor of Educational Technology in the Childhood/Early Childhood Education Department at SUNY Cortland. His research interests include the preparation of teachers for blended

learning environments, technology integration to meet the needs of diverse learners, and the integration of robotics and computational thinking with young children.

Jing Lei is a Full Professor in the Instructional Design, Development and Evaluation Department at Syracuse University. Her research interests include educational technology integration, meaningful technology use in schools, social-cultural and psychological impact of technology, teacher technology professional development, and international and comparative education.

Qiu Wang is an Associate Professor for Measurement and Research Methodology in the Department of Higher Education, Syracuse University. His research interests include educational assessment, psychometrics methods and statistical modeling.