## *Contents*

# Articles

# Relationships Between Access to Mobile Devices, Student Self-Directed Learning, and Achievement

*Scott R. Bartholomew, Ed Reeve, Raymond Veon, Wade Goodridge, Victor Lee, & Louis Nadelson*

**Abstract**

Today's students are growing up in a world of constant connectivity, instant information, and ever-changing technological advancements. The increasingly ubiquitous nature of mobile devices among K–12 students has led many to argue for and against the inclusion of these devices in K–12 classrooms. Arguments in favor cite instant access to information and collaboration with others as positive affordances that enable student self-directed learning.

In this study, 706 middle school students from 18 technology and engineering education classes worked in groups of 2–3 to complete an open-ended engineering design challenge. Students completed design portfolios and constructed prototypes in response to the design challenge. Classes were divided with some allowing access to mobile devices during the study and others not allowing access. Additionally, randomly assigned classes completed the design portfolio electronically, and others completed the portfolio on paper. Final student portfolios and products were assessed and assigned a rank order using a method of assessment called adaptive comparative judgment. Thirty student interviews were conducted as well as 6 teacher interviews. Statistical analyses between student access, portfolio type, student self-directed learning, and student achievement were conducted. Findings showed that student self-directed learning was independent of mobile device access during the study. Mobile device access was significantly correlated with higher student scores on the design portfolio, but mobile device access was independent of student scores on design products.

*Keywords:* Mobile devices, self-directed learning, middle school, technology and engineering

## Need

Today's K–12 students, sometimes called "digital natives" (Prensky, 2001), are growing up in a world connected through technology. They are expected to be part of a global society that is linked through technology and to possess skills that will enable them to excel and continue as life-long learners (Johnson, Adams, & Cummins, 2012; Prensky, 2007; Tulagan, 2013; West, 2013). Today's students often have more computing power in their personal mobile devices than their parents had during their educational years (Lenhart et al., 2015). A recent study from the Pew Research Center (Lenhart et al., 2015) found that “73% of [American] teens have access to smartphones” (p. 5) and that “92% of teens report going online daily—including 24% who say they go online ‘almost constantly’” (p. 2).

Leveraging mobile devices to positively impact student achievement and self-directed learning, as well as the potential pitfalls associated with mobile devices in the classroom, has been a topic of recent discussion (Elder, 2009; Johnson, Adams, & Haywood, 2011; Lloyd, 2010; O’Bannon & Thomas, 2015; Quillen, 2010; Schenker, 2009; Shuler, 2009). However, the possibility of utilizing mobile devices to enhance student self-directed learning has not been explored.

Self-directed learning, a process in which individuals take the initiative to diagnose their own learning needs, identify resources for learning, and then evaluate their own learning (Knowles, 1975), is becoming increasingly relevant in today’s educational landscape (Mitchell, 2014). There is a potential for mobile devices to facilitate self-directed learning. As Fahnoe and Mishra (2013) noted, opportunities for learners to be self-directed are often experienced with, and as a result of, technology.

## Purpose

The purpose of this study was to determine what effect, if any, the use of mobile devices (e.g., iPad or smartphone) had on student self-directed learning and achievement in a middle school technology and engineering education (TEE) classroom during an open-ended engineering design activity. The findings from this research may benefit school administrators, teachers, parents, and students as the ongoing debate regarding the inclusion of mobile devices in the classroom continues. On a larger scale, the purpose of the study was to inform policy and decision makers as the face of education continues to change and evolve with the rapid advancements in technology. Mobile devices are one example of a potentially educational technology—an addition to the classroom that may facilitate learning and improve performance (Januszewski & Molenda, 2008). The two research questions that guided this study were:

1. What relationship, if any, exists between middle school student access to mobile devices and student self-directed learning?
2. What relationship, if any, exists between middle school student access to mobile devices and student achievement on an open-ended design problem?

Although this study specifically looked at the influence of access to mobile devices on student self-directed learning and achievement, it should be noted that the findings of this study should not necessarily be confined to mobile devices. Mobile devices most directly offer the added benefit to students of access to information in real time, communication, and other functionalities. The findings from this study can be used to inform current thinking and inquiry regarding the place, use, and implementation of mobile devices. On a larger scale, these findings can be used as another resource in the debate surrounding personal access to the Internet, communication, and other functionalities in public schools.

## Self-Directed Learning

Self-directed learning has been identified as a key 21st century skill required for students to succeed (Fahnoe & Mishra, 2013; Partnership for 21st Century Learning, 2015; Zsiga & Webster, 2007). However, the majority of current research related to self-directed learning is about adult learners not K–12 students (Fahnoe & Mishra, 2013; Liu et al., 2014). Self-directed learning combines an understanding of what is not known with an understanding of what activities need to be undertaken in order to obtain the needed knowledge (Van Deur, 2004). Self-directed learning includes “self-managing, self-monitoring, and self-modifying capabilities [, which] . . . characterize[s] peak performers in all walks of life” (Costa & Kallick, 2004, p. 52).

Self-directed learning has been identified as positively correlated with numerous characteristics, including GPA, openness, conscientiousness, emotional stability, extraversion, optimism, career-decidedness, work drive, life satisfaction, and self-actualization (Lounsbury, Levy, Park, Gibson, & Smith, 2009). In one study, Fahnoe and Mishra (2013) examined sixth graders’ self-directed learning as it corresponded with technology use. Utilizing the Self-Directed Learning with Technology Scale (SDLTS; Teo et al., 2010), Fahnoe and Mishra (2013) reported that students in the technology-rich environment were statistically significantly more self-directed in their learning than their classmates in the traditional classroom, suggesting that technology carries with it the possibility of increasing and encouraging self-directed learning in K–12 students.

In their article “Students’ Perceptions of Self-Directed Learning and Collaborative Learning With and Without Technology,” Lee, Tsai, Chai, and Koht (2014) found “that students who reportedly engaged in SDL [self-directed learning] and CL [collaborative learning] in face-to-face contexts also engaged in these forms of learning in technology-supported contexts” (p. 425), suggesting that self-directed learning practices may occur independently of the presence of technology. Exploring the influence of technology on the self-directed learning practices of students was one goal of this study.

## Mobile Devices in K–12 Education

Literature related to mobile devices spans a variety of settings, devices, and definitions. This article focuses on mobile devices and mobile learning in K–12 classrooms and used Traxler’s (2005) definition of *mobile learning*: “any educational provision where the sole or dominant technology is a handheld or palmtop device” (p. 265). Additionally, this study utilized Kim, Holmes, and Mims’ (2005) definition for *mobile wireless technology* (or mobile devices):

"technology that provides continuous accessibility to users anytime, anywhere without using a wire or cable to connect to networks (like the internet), transmit data or communicate with others" (p. 55). For this study, the two identified definitions were combined to define mobile learning with the inclusion of mobile devices: "any educational provision where the sole or dominant technology is a handheld or palmtop device" (Traxler, 2005, p. 265) "that provides continuous accessibility to users anytime, anywhere, and without using a wire or cable to connect to networks (like the Internet), transmit data, or communicate with others" (Kim, Holmes, & Mims, 2005, p 55).

Despite the rapid increases in mobile devices, mobile learning, and educational technology opportunities, research related to mobile devices in K–12 settings is limited in scope (Cheung & Hew, 2009; Hwang & Tsai, 2011; Liu et al., 2014; Sutton, 2011; Wan, 2011). Although the impacts of mobile devices in K–12 classrooms are relatively unclear (Cheung & Hew, 2009; Hwang & Tsai, 2011; Liu et al., 2014; Sutton, 2011; Wan, 2011), there have been recent discernable efforts at implementing more "mobile friendly" policies and incorporating mobile devices into student learning experiences (Hwang & Tsai, 2011; Liu et al., 2014; Lloyd, 2010; Quillen, 2010; Schenker, 2009; Shuler, 2009). The benefits of including mobile devices in K–12 classrooms seem to center around student access to information, others, and technology (Lenhart et al., 2015; Prensky, 2007; Robledo, 2012; Shuler, 2009; West, 2013). Interestingly, Mentzer (2011) showed that access to information (i.e., the Internet) did not improve student designs in an open-ended engineering design challenge when compared with other students without access. Relatedly, Pieper and Mentzer (2013) found that students with access to the Internet during an open-ended design challenge spent significantly more time accessing information than their peers without Internet access; however, this additional time was not always productive or impactful. This study aimed to add additional insight to the question of whether or not access to mobile devices, and in turn information, will be impactful on student learning in open-ended engineering design challenge settings.

## Adaptive Comparative Judgment

In this study, the adaptive comparative judgment (ACJ) method was used to assess student product and portfolio performance. ACJ is a relatively new form of assessment, originating in the United Kingdom, and this is the first time that it has been used in a middle school research study in the United States. ACJ was developed through work by Alastair Pollitt and Richard Kimbell (see Pollitt, 2004, 2007, 2012 and Kimbell, 2007, 2012) and relies on comparisons rather than rubrics or scores for assessing student work. ACJ, based on Thurstone's law of comparative judgment (Thurstone, 1927), is a form of assessment in which judges are presented with two different artifacts of student work (in the case of this research, the judges viewed two design portfolios or two student products). Each judge is not asked to grade either of the artifacts but rather to simply make a holistic judgment about which artifact is better based on a provided rubric and their own professional opinion. This process is repeated a number of times until a rank-order is produced for the artifacts viewed by the judges.

Arguing in favor of this form of assessment, Pollitt (2004) and Kimbell (2012) point out that although current trends in education often favor rubrics, assessment of any kind ultimately involves the comparison of one thing to another.

> All judgements are relative. When we try to judge a performance against grade descriptors we are imagining or remembering other performances and comparing the new performance to them. (Pollitt, 2004, p. 6)

Following the theoretical development of the ACJ process, a grading engine was commercialized by TAG Assessment under the name *CompareAssess*. Using a complex algorithm, which has been validated repeatedly and used on thousands of student artifacts (Pollitt, 2004, 2012), *CompareAssess* combines rankings from a panel of judges to assign a final rank order to each artifact. In the *CompareAssess* engine, each artifact is compared with other artifacts by randomly assigned graders until a specified reliability requirement has been met. The reliability obtained is best understood as the judge consistency coefficient—similar to an inter-rater reliability level (Pollitt, 2015)—and this method of assessment has repeatedly demonstrated more reliability and validity than traditional methods of assessing student work (Kimbell 2012; Pollitt, 2004, 2007, 2012).

An additional point mentioned in the ACJ literature relates to the method's validity; ACJ results were compared with ranking results through traditional methods, and the resulting value of $R^2$ was 0.81, corresponding to a correlation of 0.90 (Kimbell, Wheeler, Miller, & Pollitt, 2007). These findings further suggest that the ACJ method of scoring is valid and will produce highly correlated results to traditional marking.

Interestingly, although ACJ is not widely used in the United States, the ACJ method of assessment shares some similarities to other innovative assessment techniques being piloted. Denson, Buelin, Lammi, and D'Amico (2015) recently published their work on developing a creativity assessment that makes use of an online platform for viewing and rating pieces of student work. Although this method did not use ACJ, it had other functions similar to *CompareAssess* and demonstrates a larger interest in alternative and more effective methods of assessment.

## Methodology: Mixed-Method, Quasi-Experimental Study

### Pilot Study

Following Internal Review Board approval from the participating schools and the university, a pilot study was undertaken at a local middle school with two classes (the Exploring Technology class for seventh and eighth grade students). Each class received the same instruction and completed identical design challenges. One class completed the design portfolio on paper without access to mobile devices, and the other had access to mobile devices and completed the same design portfolio using an iPad app for portfolio creation entitled *LiveAssess*. *LiveAssess* was developed concurrent to the ACJ assessment engine through the efforts of Kimbell (2007) and similarly commercialized by TAG Assessments. The purpose of the pilot was to experiment with the research process, instruments, teacher pacing guide, and design challenge. During the pilot study, the researcher took copious notes regarding minor tweaks, language changes, and areas of confusion for the participating students. These notes, in addition to student comments regarding possible improvements (as gathered through post-pilot student questionnaires), were all used to revise the study prior to full implementation.

### Research Design

Implementation of the full study took place in a large suburban school district in the western United States. This large school district is one of the 50 largest school districts in the United States with a primarily suburban middle-class population (16% free or reduced-price lunch). A total of six teachers were recruited for the study based on willingness to participate and possessing similar characteristics (teacher license level, similar years of teaching, similar classes taught, similar school facilities, and recommendation from the district TEE coordinator). Each teacher agreed to implement the study in at least two sections of the Exploring Technology class, an introductory TEE course for seventh and eighth graders. A total of 706 students were included in the study, which required five class periods (90-minute class periods every other day for 2 weeks). A total of 18 classes of the Exploring Technology course from the six teachers formed the population of the study. Two teachers used paper portfolios with their classes, and four teachers used iPads to complete the portfolios via the *LiveAssess* app. Four randomly assigned teachers (i.e., one paper-based portfolio, three iPad-based portfolio) were instructed to allow ubiquitous mobile device access, and the other two teachers (one paper-based portfolio, one iPad-based portfolio) were instructed to prohibit this access during the unit. The counter-balanced nature of the access and portfolio medium was undertaken in an effort to highlight possible problematic variables related to the dependent measures.

Teachers were trained prior to the study in a 2-hour training session conducted by the researcher during which teachers were provided with paper and electronic access to all study and training materials. Teacher compliance and fidelity to study measures and to the provided teacher script were monitored through daily observations by the researcher and by means of responses to qualitative interviews at the conclusion of the study.

Students began the study by completing a pre-study questionnaire. This questionnaire included demographic questions, inquiries regarding their use and comfort with technology from the Digital Natives Assessment Scale (DNAS; Teo, 2013), and questions adapted from the Self-Directed Learning with Technology Scale (SDLTS; Teo et al., 2010). Following the pre-study questionnaire, students received instruction related to mobile device use, digital citizenship, and the engineering design process. Students were then placed in groups of 2–3 to complete an open-ended engineering design challenge. This challenge involved the designing of a new container or dispenser for distributing pills to patients in specified quantities and at prescribed times (see similar examples in Kimbell, 2007, 2012). Students designed the product for a specific user: an elderly individual who enjoys traveling internationally.

Initially, groups of students were provided with a “handling collection” consisting of materials chosen to stimulate idea generation and creativity (e.g., zippers, ties, string, plastic, and clay). Students were also shown pictures of pill holders and containers as well as the student creations from the pilot study. Following this brainstorming activity, students returned the materials in the “handling collection” and were provided with new materials from the “modeling collection” that was used to construct a solution to the design problem. Following prompts from their teachers, students filled out a design portfolio (either on paper or electronically) throughout the design challenge. The overall progression through this activity was managed by means of a provided teacher script that instructed teachers when to prompt students to complete a portion of their portfolio and when to move to a new portion of the lesson. The design portfolio was loosely influenced by similar portfolios used in Kimbell’s research (2007, 2012) and was crafted to help the students display their progress through the design process.

Students worked on their designs and portfolios for four class periods and on the final day (Day 5), students turned in their portfolios and products and completed a post-study questionnaire. Teachers identified five students for the researcher to interview: two “high-performing” students, two “low-performing” students, and one “average-performing” student. The researcher conducted a semi-structured qualitative interview with these students and asked them questions related to self-directed learning, mobile devices, engineering design, and their experience with the study. Teachers were also interviewed and asked similar questions in an effort to further explore, clarify, and highlight the findings from the study.

**Mixed-Method Data Collection**

Following the completion of the study, all the student products were collected, and a digital picture was taken of each one, resulting in 175 images of student design products. These pictures were uploaded to the *CompareAssess* ACJ engine for later use. Each paper portfolio was also "digitized" using a scanner and iPad to record student responses from the paper portfolios into electronic versions via the *LiveAssess* app. These were also added to the *CompareAssess* engine. Data responses to the pre- and post-study questionnaires were conditioned and matched, resulting in a pre-study, post-study, and combined data set for later analysis.

Student and teacher responses to qualitative interviews were transcribed and analyzed using descriptive and thematic coding following recommendations by Saldaña (2013). In the first step of this process, the transcriptions were analyzed, and several words that described the contents of the response were identified. The second step in the process involved a second review of the transcribed responses in which the descriptive responses were analyzed for general ideas and themes. In the third step, the identified themes and ideas were synthesized into overarching themes for each response. These themes were checked for triangulation with topics relevant to the study (e.g., mobile devices, self-directed learning, and open-ended problems). Following the quantitative analyses, the resulting final qualitative themes were used to clarify, expand, and inform the general findings from the quantitative portion of the study as well as highlighting future areas of research deserving exploration. Representative phrases and illuminating remarks from the student and teacher responses were included as illustrative examples of the overall findings.

A panel of five graders was formed, which included a Technology and Engineering Education professor, an Engineering Education Professor, an Art and Design Professor, a former middle-school teacher, and a graduate student in Technology and Engineering Education. The panel of graders was trained on the *CompareAssess* software and discussed the grading procedure together prior to completing judgments. Initially, each grader was given a login to the *CompareAssess* online judging platform and was asked to grade 20–30 portfolios and student products. Following this, an additional meeting was held to ensure a unified direction in judgment. Judges were asked to complete additional judgments up to 175 judgments of portfolios and 175 judgments of products, which resulted in a reliability coefficient of $r = .943$ for students products and $r = .934$ for student portfolios. Twenty additional judgments for portfolios and 20 more judgments for products were completed by each judge, which increased the reliability to $r = .959$ for student products and $r = .972$ for student portfolios. The result was a rank-order for student products and portfolios that was added to the statistical data set for later analysis.

Prior to analyses, regression diagnostics, including linearity, homoscedasticity, normality of residuals, uncorrelated error, mean independence, and normally distributed error, were conducted to ensure that the proper assumptions of were met for the statistical tests. It was determined that each of the tests was satisfied and that the assumptions were met. Following this, all quantitative data were analyzed using a variety of statistical procedures, including *t*-test, ANOVA, ANCOVA, correlation, and regression.

## Quantitative Findings

The quantitative findings from the study were taken from three different sources: the pre-study questionnaire ($n$ = 555), the post-study questionnaire ($n$ = 458), and the matched questionnaire ($n$ = 221) containing student pre- and post-study matched responses. The large decrease in *n*-size between the pre-study questionnaire and the matched questionnaire was due mainly to student error in entering identical unique identifiers on both the pre- and the post-study questionnaires. In order to ensure comparability between the data, independent samples *t*-tests were computed comparing the pre-study data with the combined data set on the following measures to test for significant differences: pre-study SDLTS score, DNAS score, average grades, average time spent with technology, average mobile device use, and average mobile device skill. The only test that revealed a significant difference between the pre-study data set and the combined data set was for average grades, $F$ (772) = 6.13, $p$ = .023. A follow-up independent samples *t*-test, which compared the grades in TEE classes across the groups, did not return significant results ($p$ = .17). These tests demonstrate that in all tested cases, with the exception of average grades, the students in the combined data set were not significantly different from the total *n* contained in the pre-study data set. It was thus concluded that, although not equal, the combined data set is comparable, representative, and suitable for use in further data analyses

### Self-Directed Learning Findings

A simple linear regression was calculated to predict student self-directed learning (post-study questionnaire score) based on demographic variables (age, grades in all classes, grades in TEE classes, computer and mobile device access, time spent with technology, and pre-study SDLTS score). Upon initial investigation, it was shown that not all predictors were significant to student post-study SDLTS score. Non-significant factors were removed case by case until only significant factors were contained in the regression. This resulted in a significant regression equation ($F$ (2, 218) = 26.26, $p$ < .001), with an adjusted $R^2$ of .19, and two significant predictors of student score on the post-study SDLTS assessment: average mobile device skill level and computer access and use at school (see Table 1). Student post-study SDLTS score is represented by 2.94 + .40(average mobile device skill level) - .18(computer access and use at school), suggesting a positive correlation between average mobile device skill level in students and self-directed learning and a negative correlation between computer access and use at school and student self-directed learning.

**Table 1**
*Regression Equation Results for Student Demographic Information and Post-study SDLTS Score*

| Variable | Coefficient *B* | *p*-value | *t* | *r* |
|---|---|---|---|---|
| Computer access at school | -.07 | $p = .003$ | -3.02 | -.18 |
| Mobile device skill level | .29 | $p < .001$ | 6.61 | .40 |

A paired-samples *t*-test was used to determine if there was a significant difference in the student SDLTS pre- and post-study questionnaires. The results evidenced a significant difference in student pre-study ($M = 3.61$, $SD = .54$) and post-study ($M = 3.79$, SD = .57) scores, $t = 6.521$, $p < .001$, $d = -.44$, indicating that students were more self-directed following the study.

It was anticipated that student scores on the DNAS would be predictive of their post-study SDLTS scores. A correlational analysis revealed a significant correlation ($p < .001$) in the positive direction between student DNAS and student pre-study SDLTS as well as student post-study SDLTS scores (Table 2), suggesting that higher levels of "digital nativeness" among students corresponded with higher self-directed learning.

**Table 2**
*Correlation for Student DNAS Scores and Student Pre- and Post-Study SDLTS Scores*

| | Pre-SDLTS score | Post-SDLTS score |
|---|---|---|
| Pearson correlation | .40 | .31 |
| Sig. (2-tailed) | .00 | .00 |
| *n* | 221 | 221 |

Different mediums were purposely utilized for student design portfolios as part of the counter-balanced study format. In order to separate significance based solely off the difference in portfolio medium, tests were run to determine the impact of paper or electronic portfolios on student post-study SDLTS score. Utilizing an ANCOVA, with student pre-study SDLTS score as the covariate, portfolio type and student post-study SDLTS were analyzed. The resulting *p*-value was not statistically significant ($p = .132$), suggesting that student post-study SDLTS score was independent of their assigned portfolio creation medium.

Using ANCOVA statistical techniques, analyses were conducted examining the relationship between student access to mobile devices and student post-study SDLTS score, using students' pre-study SDLTS score as a covariate. The resulting value, $p = .816$, was not significant, suggesting that student scores on self-directedness in learning with technology are independent of access to mobile devices.

A simple bivariate correlation test was conducted to look at the relationship between student comfort level with open-ended design problems and post-study SDLTS score. This reflected a significant correlation ($p < .001$) in the positive direction, suggesting that higher comfort levels with open-ended design problems are correlated with higher post-study SDLTS scores.

Using a simple correlation test, the relationship between student comfort level in working with groups and student post-study SDLTS scores was found to be significant ($p < .001$) and positive, suggesting that higher comfort in working in groups was correlated with higher post-study SDLTS scores.

**Achievement Findings**

Student achievement was measured in two ways as part of this study: student rank score on their group portfolio and student rank score on their group product (created during the design challenge). Possible relationships between student final scores and other potential predictors were explored using a variety of statistical methods.

Using correlation statistical analyses, the relationships between student group portfolio score (rank) and student group product score (rank) were identified. Table 3 outlines the relationships between student portfolio rank score and demographics with several significant ($p<.05$) correlations (age, grades, time with technology, skill with mobile devices, and access to mobile devices). When compared, the relationship between student product rank score and demographics demonstrated that age was the only significant correlation ($p = .05$).

**Table 3**
*Student Demographics Measures and Student Portfolio Rank Score*

| Student portfolio rank | Spearman correlation | Sig. (2-tailed) | *n* |
|---|---|---|---|
| Student age | .16 | .02 | 221 |
| Grades in average (all classes) | .13 | .05 | 220 |
| Grades on average (TEE only) | -.02 | .83 | 221 |
| Average time using technology | .27 | .00 | 214 |
| Average mobile device use | .05 | .45 | 221 |
| Skill level with mobile devices | .15 | .02 | 221 |
| DNAS score | .12 | .08 | 221 |
| Prestudy SDLTS score | -.07 | .33 | 221 |
| Computer access (home) and use | .05 | .50 | 221 |
| Computer access (school) and use | .09 | .17 | 218 |
| Mobile device access (home) and use | .27 | .00 | 219 |
| Mobile device access (school) and use | .24 | .00 | 219 |

A set of simple correlation tests revealed that the correlation between student pre-study SDLTS score and their portfolio rank score was not significant ($r$ = -.07, $p$ = .33). The correlation between student pre-study SDLTS score and their product rank score was also not significant ($r$ = -.05, $p$ = .48).

A correlation was computed for student self-directed learning, as measured on the post-study *SDLTS*, and student rank portfolio score. A correlation was also computer for student self-directed learning, as measured on the post-study SDLTS, and student rank product score. Neither relationship returned a significant value.

Looking at correlation tests, the correlation between student DNAS scores and their product rank score was not significant ($r$ = -.04, $p$ = .54). The correlation between student DNAS scores and their portfolio rank score approached significance ($r$ = .12, $p$ = .08) but was not significant.

Using an independent samples $t$-test, the impact of portfolio type on student achievement (both portfolio and product rank scores) was analyzed. There was a significant difference in student product scores between paper ($M$ = 73.93, $SD$ = 52.22) and electronic portfolios ($M$ = 97.71, $SD$ = 49.63); $t$ (455) = -4.83, $p$ < .001. There was also a significant difference in student portfolio scores between paper ($M$ = 68.83, $SD$ = 39.46) and electronic portfolios ($M$ = 96.58, $SD$ = 53.43); $t$ (454) = -5.84, $p$ < .001. It is important to note that the scores for the portfolios and the products are rank scores, so a lower rank is deemed of higher quality than a higher rank.

Using an independent samples $t$-test, the impact of mobile devices on student achievement (both portfolio and product rank scores) was analyzed. There was a significant difference in student portfolio scores between those with access to mobile devices (*M* = 81.65, *SD* = 52.07) and those without access to mobile devices (*M* = 101.29, *SD* = 42.52); $t$ (454) = -3.62, $p < .001$. However, there was not a significant difference in student product scores between those with access to mobile devices (*M* = 90.20, *SD* = 52.82) and those without access to mobile devices (*M* = 85.60, *SD* = 48.60); $t$ (455) = .816, $p = .415$.

A one-way ANOVA was computed to assess the impact of the teacher on student achievement scores for the portfolio. The results were significant ($F$ = 37.70, $p < .001$), and LSD post hoc analyses were computed to further explore the difference between teacher groups (see Table 4).

**Table 4**
*Post Hoc Analysis of Differences in Student Product Rank by Teacher*

| Teacher (*n*, *M*, *SD*) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Teacher 1 (85, 77.99, 47.39) | | .00 | .00 | .35 | .00 | .79 |
| Teacher 2 (84, 107.17, 51.11) | | | .94 | .00 | .62 | .00 |
| Teacher 3 (69, 106.54, 48.26) | | | | .00 | .69 | .00 |
| Teacher 4 (59, 70.10, 47.20) | | | | | .00 | .46 |
| Teacher 5 (53, 102.85, 44.51) | | | | | | .00 |
| Teacher 6 (107, 76.05, 54.89) | | | | | | |

A separate one-way ANOVA was computed to assess the impact of the teacher on student achievement scores for the product, which also returned significant results ($F$ = 8.77, $p < .001$). This necessitated LSD post hoc analyses to further explore the difference between teacher groups (see Table 5). Both Tables 4 and 5 highlight significant differences in students achievement based on teacher.

**Table 5**
*Post Hoc Analysis of Differences in Student Portfolio Rank by Teacher*

| Teacher (*n*, *M*, *SD*) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Teacher 1 (84, 64.26, 48.98) | | .00 | .00 | .00 | .00 | .12 |
| Teacher 2 (84, 130.55, 44.32) | | | .00 | .00 | .00 | .00 |
| Teacher 3 (69, 85.20, 45.78) | | | | .22 | .00 | .00 |
| Teacher 4 (59, 94.58, 36.23) | | | | | .08 | .00 |
| Teacher 5 (53, 108.75, 47.83) | | | | | | .00 |
| Teacher 6 (107, 86.47, 50.56) | | | | | | |

## Qualitative Findings

The final themes emerging from the student and teacher interviews were used to triangulate, clarify, and expand the quantitative findings resulting from the analysis of the student questionnaires and final rank order of the products and portfolios. The themes, along with representative responses, are included here.

### Self-Directed Learning

Student and teacher comments related to self-directed learning revolved around the necessity of student choice for self-directed learning to occur. Two students commented on student choice in defining self-directed learning.

> [Self-directed learning is] something that you, like go and do yourself, like you are interested in it, you want to go and figure out what this thing is . . . or how something works.

> [Self-directed learning is] somebody actually choosing what they have to do and what they want to do in their education.

One teacher comment also highlighted the student initiative and choice involved with self-directed learning.

> Self-directed learning is where a student takes their own personal initiative to take the supplies that I've provided and also the knowledge that I have provided that they need, and of their own knowledge and their own supplies—based off of rules and I guess regulations, based off of our assignments or whatever—to create a learning environment where they are benefitted.

### Mobile Devices

Teachers and students responded to questions regarding the potential benefits and challenges of mobile devices in K–12 education. Their responses themed around (a) mobile devices being enablers or both positive and negative behavior, (b) mobile devices being regulated by strict rules and monitoring, and (c) classroom norms acting in opposition to mobile device integration. Examples of student comments related to the enabling nature include the following.

> [Mobile devices] help, because you can look . . . like if you want to learn something, like if you were trying to teach yourself how to play the guitar or something you could look up videos online of how to do it.

> Well it just matters on the kid pretty much. I think that [mobile devices] would help most kids, but some kids are just there to get the grade and to

> dink off with it and ruin the privileges. It would help them because like, they, oh I feel familiar with this—I know what to do, I know where to go.

> I think [mobile devices] would help some kids, but some kids would just play on them, and then, maybe look up a few things . . . I would use mine for, uh, learning because I don't really like being on social media, but I don't know about other people very well, I just see a lot of people on Instagram during lunch. So . . . I'm not really, out to using it the same way as other people.

Students identified strict rules, regulations, and monitoring as both the reality and a necessity for mobile devices in K–12 classrooms. Student comments also centered on different areas where mobile devices were allowed and other areas where mobile devices were prohibited. Example student responses include the following.

> I think [mobile devices] would . . . help, but there would have to be restrictions, 'cuz if kids were just playing on their phones, they wouldn't be learning, and they wouldn't like, be paying attention to the teacher. So they wouldn't get the grade they want on their test, and, so, that would bring grades down, but like using them would in like, effective ways in schools, would bring them up.

> In school [mobile devices] are allowed during class, if the teacher gives you permission, only if you are, like, working on an assignment or something. Um, they are allowed during lunch—private time—before and after school. Um, and like usually people just, like, use them to do, like, calculators or math, and stuff like that.

> It all depends like what class, like [mobile devices] are not allowed in like, during class but some teachers like let you use them for like certain things if you don't know, like, how to like, spell something or like draw something then you're allowed to use them.

Teachers' comments were similarly themed to the student responses in regards to the need for rules and regulations in order for mobile devices to be successful in K–12 classrooms. One example response illustrates this general consensus among teachers:

> I think that [mobile devices in K–12 classrooms] can be good in a monitored fashion, with activities like the one we did, or other experience design activities. It could be very valuable in the research and

> understanding what the actual problem is they're trying to solve and where it fits in the world of what the impact that decision or solution might have.

Regardless of permission to use mobile devices, the majority of students did not choose to use mobile devices regularly throughout the study. When asked about the reasons guiding the students' decision to use or not to use a mobile device during the study, teachers highlighted classroom norms as a potential reason for students not using mobile device: Students and teachers were accustomed to a restriction on device use in class that prevailed despite permission to use the devices. One teacher remarked,

> I had a couple kids looking on the iPad on the Internet. Honestly I was surprised that when we opened it up to the mobile devices more students didn't have their cell phones out. Most of them were just looking for images in [one] of the pill bottle folder things. But I was surprised at, I guess, the lack of using that device. Maybe it's because they're not used to using it in my classroom—I really don't know. The only thing I can think of is because it's the rule that you don't have your cell phone out in my class, I kind of felt like that was the norm.

## Summary of Findings

For the middle-school students in this study, self-directed learning appeared to be related to student and environmental characteristics rather than access to specific technology tools. When analyzed, student self-directed learning was independent and even negatively correlated with access to some technology tools (e.g., mobile devices and computers), and student self-directedness in learning scores were independent of student portfolio type (paper vs. electronic). Taken together, these findings suggest that technology tools in and of themselves may not correspond with an increase in student self-directed learning and, in some cases, may be detrimental to student self-directed learning. These findings appear to align with Mentzer's (2011) research, which also concluded that access to information (i.e., the Internet via computers) did not significantly improve student designs.

Unlike technology tools, a variety of specific student and classroom environment characteristics did demonstrate significant relationships with student self-directed learning. Student characteristics that corresponded with higher levels of self-directedness in learners were: average skill in using mobile devices, higher "digital nativeness" scores, student familiarity with open-ended design problems, and student comfort level in working in groups. Responses in teacher interviews seem to concur with this. Teachers discussed how they perceived self-directed learning to be a product of external conditions such as: the presence of an open-ended problem, a task involving group work, or other classroom-environmental factors.

Student achievement was identified through two separate student scores: student portfolio scores and student product scores. In qualitative interviews, the teachers and students were in agreement that mobile devices had the potential to improve students' achievement if used correctly.

A key finding is that teachers and portfolio medium (paper vs. electronic) were the most significant factors in student achievement. Students completing portfolios on paper produced significantly better portfolios and products than their counterparts who completed electronic portfolios. Despite the fact that all teachers in the study were comparable, there were significant differences in the final grades received by the students of each teacher, with the students of one teacher (Teacher 6) scoring significantly higher than the other students in the study. This teacher was assigned to complete the portfolios on paper, which may be a confounding factor resulting in the paper portfolios being ranked much higher than the electronic portfolios and this teacher's students outperforming the others. These findings align with other research demonstrating the significant impact of a teacher on their students above and beyond other factors (Darling-Hammond, 2000).

**Student Portfolios**

Student access to mobile devices was significantly correlated with higher study scores on the design portfolio. Average time spent with technology, student age, mobile device skill level, and mobile device access at home and school were also significantly correlated with higher scores. Student pre-study SDLTS and student post-study SDLTS scores were both independent of student portfolio score rank—an important finding suggesting that self-directed learning may not be indicative of student achievement, ability, and skill with the engineering design process despite its identification as a key skill for 21st century learners (Partnership for 21st Century Learning, 2015).

**Student Products**

Unlike the portfolio scores, the only significant correlation found between student product scores, aside from teacher and portfolio type, was student age. Older students tended to receive better scores on their design products. Student portfolio scores were not significantly correlated with pre- or post-study SDLTS, pre-study DNAS score, or access to mobile devices. Interestingly, the two teachers with the overall top-performing students (Teachers 4 and 6) had the youngest average students in their participating classes across the study. This emphasizes the strength of the impact made on students' achievement by their teacher.

**Other Observations**

Of particular interest, the researcher noticed that although many students were given access to mobile devices, students rarely used mobile devices during the product creation or the portfolio creation. Teachers echoed this sentiment during interviews and provided several conjectures for lack of mobile device usage, including: lack of need for mobile devices, the competition between computers and mobile devices, and classroom norms. Although students cited specific benefits of mobile devices in the interviews, the majority (65.4%) of students who were given access to mobile devices during the study reported using mobile devices less than 30 minutes during class over the course of the entire study (over 360 minutes of class time). In the interviews, students mentioned that this activity was the "wrong type of problem" for using a mobile device. When asked for clarification, students commented that they were unsure how to use a mobile device to help them with an open-ended problem and were most comfortable using their mobile devices to answer factual single-answer problems. Student and teacher interview responses also themed around classroom norms: Although mobile devices were allowed, the previously established classroom norm (no mobile devices allowed) appeared to be highly influential on students' choices regarding mobile device use.

## Further Research and Analysis

Using both the quantitative and qualitative findings from this study, the following recommendations and areas for further research and analysis were identified.

### Self-Directed Learning

As noted, mobile devices did not make a significant impact on student self-directed learning as measured by the SDLTS on the pre- or post-study questionnaires. However, several other student and classroom environment characteristics were positively correlated with self-directed learning in a significant way, including student skill in using mobile devices and student "digital nativeness." This suggests that teachers and schools should emphasize student skills in using and interacting with technology as a means of improving self-directed learning. If students can more effectively interact with different technologies (e.g., mobile devices, tablets, and computers), their opportunities and abilities for self-directed learning may also increase.

### Mobile Devices

Although mobile devices did not significantly impact student self-directed learning in this study, mobile devices did correlate significantly with higher student achievement on the design portfolio. During student interviews, a theme that emerged with relation to mobile devices was the need for direct instruction regarding how, when, and where students should use mobile devices. Teachers can work through explicit instruction so those students understand how to use mobile devices and so that positive and appropriate uses of mobile devices become the new norm for their classroom.

**Further Research**

Additional relationships between student mobile device access and factors outside of self-directed learning or achievement would shed further light on the debate over mobile devices in the classroom. The findings from this study are limited in scope to a relatively suburban, middle-class, homogeneous population within specific grade levels (seventh and eighth grade). Other research with different population groups, ages, or different locations could shed additional light on and provide valuable comparisons for the findings of this study. As teacher impact was highly significant in this study, it is recommended that additional studies of varying research designs be undertaken to explore supplementary data that examine specifically the impacts of teacher influence. Additionally, student gender was not collected during this study—this has been identified in other studies as significant (Reio & Davis, 2005) and should be taken into account in future research efforts around student self-directed learning.

## Implications

Granting access to mobile devices in middle school TEE classrooms during a STEM activity appears to have the potential to transform and improve student educational experiences. Although student self-directed learning was not significantly impacted in this study by access to mobile devices, aspects of student achievement showed positive correlations with access to mobile devices. In order for mobile devices to be impactful, teachers and students will need to work together to change the classroom norms relating to mobile device use, and teachers will need to model appropriate and effective mobile device use for their students.

In this study, student self-directed learning correlated more closely with student and classroom characteristics than it did with access to technology tools, suggesting a possible shift in the debate surrounding mobile device inclusion in classrooms from the actual tools to the learner and classroom characteristics. As previously shown (Darling-Hammond, 2000), the impact of a teacher on student achievement is significant: Students' final portfolio and product scores were more directly related to their teacher than any other variable. Focus on effective teacher habit identification and training should take precedence over technology tools and other classroom add-ons.

## References

Costa, A. L., & Kallick, B. (2004). Launching self-directed learners. *Educational Leadership*, *62*(1), 51–55.

Cheung, W. S., & Hew, K. F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology*, *25*(2), 153–183. doi:10.14742/ajet.1148

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, *8*. doi:10.14507/epaa.v8n1.2000

Denson, C. D., Buelin, J. K., Lammi, M. D., & D'Amico, S. (2015). Developing instrumentation for assessing creativity in engineering design. *Journal of Technology Education*, *27*(1), 23–40. doi:10.21061/jte.v27i1.a.2

Elder, J. (2009, October 29). OMG! Teachers say texting can be good for teens. *Sauk Valley Media*. Retrieved from http://www.saukvalley.com/2009/10/29/omg-teachers-say-texting-can-be-good-for-teens/ap5xt6y/

Fahnoe, C. & Mishra, P. (2013). Do 21st century learning environments support self-directed learning? Middle school students' response to an intentionally designed learning environment. In R. McBride & M. Searson (Eds.), *Proceedings of SITE 2013—Society for Information Technology & Teacher Education International Conference 2013* (pp. 3131–3139). Chesapeake, VA: Association for the Advancement of Computing in Education.

Hwang, G.-J. & Tsai, C.-C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *42*(4), E65–E70. doi:10.1111/j.1467-8535.2011.01183.x

Januszewski, A., & Molenda, M. (Eds.). (2008). *Educational technology: A definition with commentary*. New York, NY: Erlbaum.

Johnson, L., Adams, S., & Cummins, M. (2012). *NMC horizon report: 2012 higher education edition*. Austin, TX: New Media Consortium. Retrieved from https://www.nmc.org/pdf/2012-horizon-report-HE.pdf

Johnson, L., Adams, S., & Haywood, K. (2011). *NMC horizon report: 2011 K-12 edition.* Austin, Texas: New Media Consortium. Retrieved from https://www.nmc.org/sites/default/files/pubs/1316810422/2011-Horizon-Report-K12.pdf

Kim, S. H., Holmes, K., & Mims, C. (2005). Mobile wireless technology use and implementation: Opening a dialogue on the new technologies in education. *Techtrends: Linking Research and Practice to Improve Learning*, *49*(3), 54–64. doi:10.1007/BF02763647

Kimbell, R. (2007). E-assessment in project e-scape. *Design and Technology Education: An International Journal*, *12*(2), 66–76. Retrieved from https://ojs.lboro.ac.uk/DATE/article/view/Journal_12.2_0707_RES6

Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education 22*(2), 135–155. doi:10.1007/s10798-011-9190-4

Kimbell, R., Wheeler, T., Miller, S., & Pollitt, A. (2007). *E-scape portfolio assessment phase 2 report*. London, England: Technology Education Research Unit, Goldsmiths University of London.

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. Englewood Cliffs, NJ: Cambridge Adult Education.

Lee, K., Tsai, P.-S., Chait, C. S., & Koht, J. H. L. (2014). Students' perceptions of self-directed learning and collaborative learning with and without technology. *Journal of Computer Assisted Learning*, *30*(5), 425–437. doi:10.1111/jcal.12055

Lenhart, A., Duggan, M., Perrin, A., Stepler, R., Rainie, L., & Parker, K. (2015). *Teens, social media & technology overview 2015: Smartphones facilitate shifts in communication landscape for teens.* Washington, DC: Pew Research Center. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/04/PI_TeensandTech_Update2015_0409151.pdf

Liu, M., Scordino, R., Geurtz, R, Navarrete, C., Ko, Y., & Lim, M. (2014). A look at research on mobile learning in K–12 education from 2007 to the present. *Journal of Research on Technology in Education*, *46*(4), 325–372. doi:10.1080/15391523.2014.925681

Lloyd, J. R. (2010, August 16). Cell phones head to class. *San Antonio Express News*. Retrieved from http://www.mysanantonio.com/news/education/article/Cell-phones-head-to-class-780555.php

Lounsbury, J. W., Levy, J. J., Park, S.-H., Gibson, L. W., & Smith, R. (2009). An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences*, *19*(4), 411–418. doi:10.1016/j.lindif.2009.03.001

Mentzer, N. (2011). *Engineering design thinking and information gathering final report*. Retrieved from http://ncete.org/flash/pdfs/Mentzer%20Final%20Report.pdf

Mitchell, L. (2014, March 14). Rise of the YouTube makeup artist: Nearly half of women use online tutorials. *Express*. Retrieved from: http://www.express.co.uk/life-style/style/464633/Pixiwoo-Lauren-Curtis-Frozen-Nearly-half-of-women-copy-popular-online-makeup-tutorials

O'Bannon, B. W. & Thomas, K. M. (2015). Mobile phones in the classroom: Preservice teachers answer the call. *Computers & Education*, *85*, 110–122. doi:10.1016/j.compedu.2015.02.010

Partnership for 21st Century Learning. (2015). P21 framework definitions. Retrieved from http://www.p21.org/storage/documents/docs/P21_Framework_Definitions_New_Logo_2015.pdf

Pieper, J. & Mentzer, N. (2013). High school students' use of paper-based and Internet-based information sources in the engineering design process. *Journal of Technology Education*, *24*(2), 78–95. doi:10.21061/jte.v24i2.a.6

Pollitt, A. (with Elliott, G., & Ahmed, A.). (2004, June). *Let's stop marking exams*. Paper presented at the annual conference of the International Association for Educational Assessment, Philadelphia, PA.

Pollitt, A. (2007). Grading students' work. In Kimbell, R., Wheeler, T., Miller, S., and Pollitt, A., *E-scape portfolio assessment phase 2 report* (pp. ?–?). London, England: Technology Education Research Unit, Goldsmiths University of London.

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles*, *Policy & Practice*, *19*(3), 281–300. doi:10.1080/0969594X.2012.665354

Prensky, M. (2001). Digital natives, digital immigrants: Part 1. *On the Horizon*, *9*(5), 1, 3–6. doi:10.1108/10748120110424816

Prensky, M. (2007). Changing paradigms. *Educational Technology*, *47*(4), 64.

Quillen, I. (2010, October 15). Schools open doors to students' mobile devices. *Education Week: Digital Directions.* Retrieved from http://www.edweek.org/dd/articles/2010/10/20/01mobile.h04.html

Reio, T. G., Jr., & Davis, W. (2005). Age and gender differences in self-directed learning readiness: A developmental perspective. *International Journal of Self-Directed Learning*, *2*(1), 40–49. Retrieved from https://docs.wixstatic.com/ugd/dfdeaf_aabb3036f5594ce7ad25ca4edaa6264a.pdf

Robledo, S. J. (2012). *Mobile devices for learning: What you need to know*. San Rafael, CA: Edutopia. Retrieved from http://www.edutopia.org/pdfs/guides/edutopia-mobile-learning-guide.pdf

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Schenker, L. (2009, April 19). Schools embrace txt msg. *The Salt Lake Tribune*. Retrieved from http://archive.sltrib.com/story.php?ref=/news/ci_12179870

Shuler, C. (2009). *Pockets of potential: Using mobile technologies to promote children's learning.* New York, NY: The Joan Ganz Cooney Center at Sesame Workshop. Retrieved from http://www.joanganzcooneycenter.org/wp-content/uploads/2010/03/pockets_of_potential_1_.pdf

Sutton, B. B. (2011, August 21). Mobile devices in the classroom: Stem stuff [Web log post]. Retrieved from http://www.edutopia.org/groups/science-technology-engineering-mathematics-education/70218

Teo, T. (2013). An initial development and validation of a Digital Natives Assessment Scale (DNAS). *Computers & Education, 67*, 51-57.

Teo, T., Tan, S. C., Lee, C. B., Chai, C. S., Koh, J. H. L., Chen, W. L., & Cheah, H. M. (2010). The self-directed learning with technology scale (SDLTS) for young students: An initial development and validation. *Computers & Education*, *55*(4), 1764–1771. doi:10.1016/j.compedu.2010.08.001

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. doi:10.1037/h0070288

Traxler, J. (2005). Defining mobile learning. In P. Isaías, C. Borg, P. Kommers, & P. Bonanno (Eds.), *Proceedings of the 2005 IADIS International Conference on Mobile Learning* (pp. 261–266). Qwara, Malta: IADIS Press.

Tulagan, B. (2013). *Meet Generation Z: The second generation within the giant "Millennial" cohort* [White paper]. Retrieved from http://rainmakerthinking.com/assets/uploads/2013/10/Gen-Z-Whitepaper.pdf

Van Deur, P. (2004). Gifted primary students' knowledge of self-directed learning. *International Education Journal*, *4*(4), 64–74. Retrieved from http://www.iejcomparative.org/data/volumes/v4n4.pdf

Wan, Y. (2011, November 4). *Mobile learning* [Memorandum]. Chicago, IL: Regional Educational Laboratory (REL) Midwest at American Institutes for Research. Retrieved from ftp://ftp.ode.state.oh.us/geodoc/Ohio_Digital_Learning/Mobile%20Learning-ed_Final.pdf

West, D. M. (2013). *Mobile learning: Transforming education, engaging students, and improving outcomes*. Washington, DC: Center for Technology Innovation at Brookings. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/BrookingsMobileLearning_Final.pdf

Zsiga, P. L., & Webster, M. (2007). Why should secondary educators be interested in self-directed learning? *International Journal of Self-Directed Learning*, *4*(2), 58–68. Retrieved from https://docs.wixstatic.com/ugd/dfdeaf_2e28e194adcc4f109eb796db14d6b134.pdf

## About the Authors

**Scott R. Bartholomew** (sbartho@purdue.edu) is Assistant Professor of Engineering/Technology Teacher Education at Purdue University.

**Ed Reeve** (ed.reeve@usu.edu) is Professor of Technology & Engineering Education at Utah State University.

**Raymond Veon** (raymond.veon@usu.edu) is Assistant Dean for Arts Education at Utah State University.

**Wade Goodridge** (wade.goodridge@usu.edu) is Assistant Professor of Engineering Education at Utah State University.

**Victor Lee** (victor.lee@usu.edu) is Associate Professor of Instructional Technology and Learning Sciences at Utah State University.

**Louis Nadelson** (lnadelson@coloradomesa.edu) is Director of Sponsored Programs and Academic Research at Colorado Mesa University.

# Educational Complexity and Professional Development: Teachers' Need for Metacognitive Awareness

*Andrew J. Hughes*

**Abstract**

The study was designed to investigate technology and engineering teachers' metacognitive awareness during specific established teacher practices. The study had a sample size of 18. There were six participants in three groups. Group 1 consisted of teachers that actively participated in Transforming Teaching through Implementing Inquiry (T2I2) professional development program. Group 2 consisted of teachers that were selected for but did not actively participate in T2I2 professional development system. Group 3 consisted of teachers that completed the National Board for Professional Teaching Standards professional development program. To measure the metacognitive awareness of each group, a semi-structured open-ended interview was used. The interviews were analyzed by two independent coders using a coding rubric. The coded interviews established the phenomenological metacognitive awareness of each group.

*Keywords:* Professional development, metacognitive awareness, technology and engineering education

*Metacognitive awareness* is the ability to recognize and regulate one's own thinking in real time. Metacognitive awareness is the term used to describe an individual's ability to detail their knowledge and regulation of cognition (Schraw & Dennison, 1994). Knowledge of cognition and regulation of cognition are two predominant components of metacognition. Examples of metacognitive subcomponents under knowledge of cognition include declarative, procedural, and conditional knowledge, and examples of metacognitive subcomponents under regulation of cognition include planning, monitoring, organizing (information management), debugging, and evaluating (Schraw, 2001; Schraw & Dennison, 1994).

The term *declarative knowledge* refers to the knowledge that a person has about their cognitive strategies, skills, and abilities (Schraw, 2001). A person with declarative knowledge knows what impacts their learning and the learning of others and what they do and do not know. A person with declarative knowledge knows strategies that can be used to increase performance for completing tasks. The term *procedural knowledge* refers to a person's knowledge about how to use strategies and techniques to increase performance and accomplish cognitive tasks (Schraw, 2001). A person with procedural knowledge will complete tasks by sequencing known strategies. The term *conditional knowledge* refers to the knowledge that a person has regarding when and why to use strategies for accomplishing tasks (Schraw, 2001). A person with conditional knowledge knows when and why to present an idea and to use strategies for completing tasks. A person with conditional knowledge can rationalize the use of specific strategies and appropriately use strategies based on the scenario.

The term *planning* is used to describe a person's ability to select appropriate strategies, set goals, and allocate resources (Schraw, 2001). Planning relates to a person's utilization of planning strategies, goal setting, and resources related to accomplishing tasks. The *organizing* subcomponent relates to the information management sub-process (Pucheu, 2008). Organizing is the use of cognitive strategies and techniques to manage information (Pucheu, 2008). *Information management* is the active process of organizing, elaborating, summarizing, and selectively focusing on important information for mental restructuring due to cognitive dissonance (Pucheu, 2008). During *monitoring*, a person assesses their cognition and strategy effectiveness (Schraw, 2001). When teachers are monitoring, they add the assessment of students' thinking through verbal and nonverbal feedback to determine their own effectiveness. During the process of *debugging* a person uses strategies to identify and correct errors and assumptions about tasks and implemented strategies (Pucheu, 2008). The subcomponent of *evaluating* is the post hoc analysis of performance and strategy effectiveness (Pucheu, 2008).

The varying complexity and duration of problems that teachers experience indicates their need for metacognition in terms of improved regulation of cognition (Hartman, 2001; Kramarski & Michalsky, 2009) and knowledge of cognition (Schraw, 2001; Wilson & Bai, 2010). Teachers that lack an awareness of their own cognitive abilities will have difficulty adapting in the constantly evolving educational environment (Kramarski & Michalsky, 2009; Lin, Schwartz, & Hatano, 2005). Prior research has established a link between teacher's metacognitive skills and the effectiveness of their teacher practices (Georghiades, 2004; Gourgey, 1998; Hartman, 2001). Metacognitive awareness is also foundational in a person's ability to learn (Ertmer & Newby, 1996). Teachers with higher levels of metacognitive awareness have improved learning capability and the ability to translate learning from professional development into classroom practices (Bransford, Brown, & Cocking, 2000; Ertmer & Newby, 1996; Pucheu, 2008).

According to Bybee and Loucks-Horsley (2000), "professional development will provide the opportunities for technology teachers and other educators to learn what they need to know and be able to do as they assist students" (p. 32) with learning, but only if teachers have the "cognitive self-awareness necessary for the kinds of metacognitive capabilities required to transfer professional development training into effective classroom practices (Bransford, Brown & Cocking, 1999; Graber, 1998; Palincsar & Brown, 1984)" (Pucheu, 2008, p. 7). The term *professional development* (PD) refers to teachers' improvement or growth of skills and knowledge, primarily with the aim of improving student achievement (Guskey, 2003; Loucks-Horsley, Hewson, Love, & Stiles, 1998). The literature presents numerous characteristics deemed essential for increasing the effectiveness of PD (Guskey, 2003; Mundry, 2007; Smylie, Allensworth, Greenberg, Harris, & Luppescu, 2001; WestEd, 2000). Gusky (2003) pointed out that not all the literature agrees on specific characteristics of effective professional development. Mundry (2007) indicated that even without "an empirically-based consensus of what constitutes effective professional development (Guskey, 2003; Whitehurst, 2002) there is a knowledge base about learning to guide the design and implementation of teacher learning programs (Elmore, 2002; Loucks-Horsley, Love, Stiles, Mundry & Hewson, 2003)" (p. 1). The ability of professional development designers to determine teachers' current level of metacognitive awareness and incorporate the further development of metacognitive awareness early on and throughout PD may impact teachers' learning ability (Ertmer & Newby, 1996; Prytula, 2012), development of pedagogical knowledge (Kramarski & Michalsky, 2009), and ability to transfer learned content back into their classrooms (Pucheu, 2008).

The research presented in this article aims to describe three different groups of teachers' metacognitive awareness while performing established teacher practices. To describe each group's metacognitive awareness, a semi-structured open-ended interview was used. The interview was designed to gather a perspective of participant's metacognitive awareness during specific teacher practices. Participants were divide into groups based on their participation in one of two different PD programs. Even though the two PD programs offered metacognitive experiences, the development of metacognitive awareness was not a focal point of either program. The findings presented in this article suggest that each group's prior level of metacognitive awareness was a factor in their successful completion of the PD program. Additionally, the findings suggest that each group's metacognitive awareness related to established teacher practices was a factor in their ability to manage educational complexity by adapting these practices.

## Background

This study was conducted in the context of a PD system called Transforming Teaching through Implementing Inquiry (T2I2), a project funded for 4 years by the National Science Foundation.[1] The project started in fall 2011 with the development of a highly interactive cyberinfrastructure system for delivering research-based PD. The PD was designed for secondary technology and engineering teachers in Grades 6–12. There were five primary goals that the designers of T2I2 attempted to accomplish. The first goal for the T2I2 program was increasing the participating teachers' ability to manage, monitor, adjust, and contribute in the learning environment. The second goal was to increase

[1] For more information about the creation of the T2I2 program, see Ernst, Clark, DeLuca, and Bottomley (2013).

teachers' understanding of engineering design concepts and the ability to effectively teach these concepts. The third goal was to increase the teachers understanding of and ability to address student learning needs. The fourth goal was increasing teachers' instructional abilities with the use of self-assessment. The final goal was to promote technology and engineering teachers' attainment of National Board Certification by aligning T2I2 with National Boards for Professional Teaching Standards (NBPTS) in three key ways: (a) the primary goals of T2I2 align with the overall goals of NBPTS, (b) the 17 learning objects of T2I2 were aligned with the 13 Career and Technical Education standards within NBPTS, and (c) T2I2 aligns with NBPTS by using shorter versions of the same teacher artifacts used by NBPTS.

Although the process of completing both the T2I2 and the NBPTS PD programs involves numerous metacognitive experiences, the importance of these experiences in developing metacognitive awareness is not communicated to participants or identified as the primary focus of either program. Because metacognitive awareness is involved throughout these experiences, it stands to reason that a teacher with already high levels of metacognitive awareness would have an easier time completing either PD program.

## Rationale

The purpose of this research was to understand technology and engineering teachers' level of metacognitive awareness in comparison to their participation and completion of either the T2I2 or the NBPTS PD program. This study was informed by research design, metacognitive, and PD literature. The literature indicated that metacognitive research often focuses on students' thinking and regulation because of the belief that metacognitive awareness helps students become better, more self-regulated learners (Schraw, 2001). However, the focus of metacognitive research has been shifting from the students to teachers due to the belief that teachers who lack metacognitive awareness are unable to help students develop their metacognitive awareness (Kramarski & Michalsky, 2009; Prytula, 2012; Pucheu, 2008). The belief that metacognitively aware teachers can help students develop their metacognitive awareness has prompted interest in PD with varying levels of focus on metacognitive awareness (Prytula, 2012; Pucheu, 2008).

A phenomenological approach was selected for this study to describe participant's metacognitive awareness related to their teacher practices (Creswell, 2007). As a phenomenological study, the research design used qualitative, semi-structured, open-ended interviews to help understand each group's metacognitive awareness during established teacher practices (Creswell, 2007; Denzin & Lincoln, 1994). The metacognitive awareness interview was designed to gather a more complete perspective of the participant's metacognitive awareness (Bryman, 2006; Creswell, 2007; Denzin & Lincoln, 1994). The semi-structured and open-ended characteristics of the interview enabled participants to provide an uninfluenced depth to their responses and promoted emergence of themes and patterns stated by each group (Akturk & Sahin, 2011; Creswell, 2007; Denzin & Lincoln, 1994). Interviewing with broad open-ended questions to investigate metacognition was supported by the literature (Akturk & Sahin, 2011; Prytula, 2012).

Questions from the metacognitive awareness interview asked the participants to detail their thinking during cognitive tasks including planning, monitoring, organizing, information management, debugging, and evaluating. Participants' ability to describe their mental phenomenon was used to indicate a level of metacognitive awareness (Georghiades, 2004, p. 374). Literature supported listening to the interviewees with as few interruptions to their responses as possible (Creswell, 2007; Denzin & Lincoln, 1994). The findings were used to provide a detailed description of the three groups' metacognition during common teacher practices.

## Method

### Instrumentation

The metacognitive awareness interview (see Table 1) was modeled based on the components and subcomponents of the Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994). The metacognitive awareness interview consisted of 11 questions, including three parts to Question 4 and two parts to Question 8. The metacognitive awareness interview questions were based on the regulation of cognition subcomponents from the MAI. Additional questions were infrequently used to guide the interviewee if the interviewer felt that they were straying from the focus of the question. The interviews were audio recorded and then transcribed for analysis.

The transcribed interviews were analyzed by two trained coders using a coding rubric. The coding rubric was generated based on the metacognitive awareness subcomponent definitions from the MAI literature. The metacognitive awareness subcomponent definitions are the descriptions of evaluated items listed in the rubric. The coding rubric was organized according to the questions in the metacognitive awareness interview, the components of metacognitive awareness, and the levels of awareness.

**Table 1**
*Metacognitive Awareness Interview*

| Question number | Interview question |
|---|---|
| 1 | Describe a method you use for planning instruction. |
| 2 | Describe a method you use for planning assessment. |
| 3 | Describe how you organize your instruction. |
| 4.1 | Remembering that there are multiple types of information; describe how you inwardly interpret information? |
| 4.2 | How do you inwardly interpret unspoken information from your students? |
| 4.3 | How do you inwardly interpret a new technique learned from another teacher or a professional development? |
| 5 | Describe how you monitor during instruction. |
| 6 | Describe how you monitor during assessment. |
| 7 | Describe how you adjust your teaching during a lesson. |
| 8.1 | Do you self-evaluate after the instructional process? |
| 8.2 | Describe how you self-evaluate after the instructional process. |

## Participants

The participants in this study were divided into three groups: (Group 1) teachers who actively participated and completed the T2I2 system, (Group 2) teachers who had been selected for but did not participate, completing less than 11% of the T2I2 system, and (Group 3) teachers who had received National Board Certification in Career and Technical Education (CTE) from the NBPTS program. This study included technology and engineering teachers from three states: Illinois, North Carolina, and Virginia. Participants from these three states applied and were randomly selected for participation in the T2I2 pilot Years 1 (2012–2013) or 2 (2013–2014). Teachers for Group 3 were also identified from these three states. The participants in Group 3 were identified through the NBPTS's website. All National Board Certified Teachers, their states, counties, certification areas, and certification expiration dates were listed. Teachers on the list were filtered by state and by CTE certification. Participants in Group 3 were identified as possible participants only if they had a currently valid NBPTS certification in CTE and were currently teaching technology and engineering education in one of the three states. A total of 73 state-certified technology and engineering teachers were initially identified for possible participation in this study; each teacher was connected to either the T2I2 or the NBPTS PD program. In order to have equal group sizes, 10 teachers from each group where randomly selected to participate.

The 30 teachers received an email explaining the study and requesting their participation. A total of 18 teachers, six from each group, responded with interest in participating with the study (Table 2). After participants made an informed decision to participate, each was assigned a unique identifying number. The participant's interview recording and interview transcript were encrypted with the unique number. The participants were sent an email in which they were asked to specify a phone number, date, and, time for the interview. Interviews were conducted by phone at the specified date and time.

**Table 2**
*Participant Group Demographics*

| | Variables | | | | | | |
|---|---|---|---|---|---|---|---|
| | Gender *n* (%) | | Experience | | | Grade level taught *n* (%) | |
| Group | Male | Female | *n* | Mean (years) | *SD* | Middle school | High school |
| 1 | 5 (83.3%) | 1 (16.6%) | 6 | 20 | 11 | 3 (50%) | 3 (50%) |
| 2 | 4 (66.7%) | 2 (33.3%) | 6 | 17.3 | 8.5 | 1 (16.7%) | 5 (83.3%) |
| 3 | 3 (50%) | 3 (50%) | 6 | 21.5 | 8.2 | 3 (50%) | 3 (50%) |
| Combined | 12 (66.6%) | 6 (33.3%) | 18 | 19.6 | 8.9 | 7 (38.9%) | 11 (61.1%) |

**Procedure**

The study was conducted over a 16-week period during fall 2014. The metacognitive awareness interview was used to gather a thorough perspective of each participant's metacognitive awareness during common teaching practices. In an attempt to build rapport with the participant, the interviewer began the interview by asking questions about the participant's background and experience related to teaching. Additional guiding questions were used at times to help the participant provide sufficient detail regarding aspects of their metacognitive awareness. The recorded interviews were transcribed and later coded by two independent and trained coders using a coding rubric (Table 3).

The two coders were selected based on their experience in teaching and qualitative research. Coder 1 had taught for six years at nearly all levels K–16. Coder 1 had also been involved numerous times in the collection and analysis of qualitative research data. Coder 2 had spent 33 years teaching at the elementary and middle school level. Coder 2 had frequently performed qualitative data collection and analysis during their career. Prior to coding, personal and identifying information were removed from the transcripts. To aid the coding process, each transcript was bracketed into sections based on the interview questions. The coders were not informed about any characteristics of the participant, including their assigned group. Each transcript was then axial coded using the coding rubric (Creswell, 2007).

The coders had to be trained regarding axial coding. *Axial coding* is the process of coding using contextual themes to encompass a phenomenon like metacognitive awareness. The training consisted of the coders reading, discussing, and rationalizing their codes based on the rubric for two transcripts. The coders completed all coding with the trainer present. During the training, the coders would frequently ask the trainer about the appropriate code to assign for a transcript section. The trainer would encourage the use of the coding rubric. The trainer would ask the coders to compare the transcript section to the corresponding descriptors in the coding rubric and to indicate the appropriate code from the rubric based on the content of that section. The coders independently used the rubric to identify each participant's level of knowledge and regulation of cognition based on their answers to the interview questions.

**Table 3**
*Metacognitive Awareness Interview Coding Rubric*

| Question focus: Sub-component | Component of meta-cognitive awareness | High level of meta-cognitive awareness (5) | High to medium level of meta-cognitive awareness (4) | Medium level of meta-cognitive awareness (3) | Medium to low level of meta-cognitive awareness (2) | Low level of metacognitive Awareness (1) |
|---|---|---|---|---|---|---|
| Declarative, procedural and conditional knowledge: Questions 1, 2, 3, 4, 5, 6, 7, and 8 | Knowledge | The participant describes a strategy; how to use the strategy, why the strategy was used in cognitive terms, and how they knew that was the strategy to use in cognitive terms. | The participant describes 3 of the 4 listed in the column High Level of Meta-cognitive Awareness. | The participant describes 2 of the 4 listed in the column High Level of Meta-cognitive Awareness. | The participant describes 1 of the 4 listed in the column High Level of Meta-cognitive Awareness. | The participant does not describe any of the 4 listed in the column High Level of Meta-cognitive Awareness. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Planning Questions 1 and 2 | Regulation | The participant describes planning, goal setting, and allocation of resources | The participant describes 2 of the items in detail and 1 item generally in the column High Level of Meta-cognitive Awareness. | The participant describes 2 of the 3 listed in the column High Level of Meta-cognitive Awareness. | The participant describes 1 of the items in detail and 2 items generally in the column High Level of Meta-cognitive Awareness. | The participant describes 1 of the 3 listed in the column High Level of Meta-cognitive Awareness. |
| Organization Question 3 | Regulation | The participant describes the implement-ation of techniques based on an under-standing of cognition for the purpose of organ-ization. | The participant describes the implement-ation of techniques based on an under-standing of cognition for the purpose of organ-ization. | The participant describes the implement-ation of techniques for the purpose of organ-ization. | The participant describes generally the implement-ation of techniques for the purpose of organ-ization | The participant does not describe the implement-ation of techniques for the purpose of organ-ization. |
| Information management Question 4 | Regulation | The participant describes their cognitive organ-ization, elaboration, summar-ization, and selective focus on important information . | The participant describes 3 of the 4 items listed in the column High Level of Meta-cognitive Awareness. | The participant describes 2 of the 4 listed in the column High Level of Meta-cognitive Awareness. | The participant describes 1 of the 4 listed in the column High Level of Meta-cognitive Awareness. | The participant does not describe any of the 4 listed in the column High Level of Meta-cognitive Awareness. |
| Monitoring Questions 5 and 6 | Regulation | The participant describes cognitive assessment of themselves, someone else, and | The participant describes 2 of the items in detail and 1 item generally in the column High Level of Meta- | The participant describes 2 of the 3 listed in the column High Level of Meta-cognitive Awareness. | The participant describes 1 of the items in detail and 2 items generally in the column High Level of Meta- | The participant describes 1 of the 3 listed in the column High Level of Meta-cognitive Awareness. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | their strategy | cognitive Awareness | | cognitive Awareness. | |
| Debugging Question 7 | Regulation | The participant describes a strategy used to correct per-formance errors and assumption s they made about a task or strategy used. | The participant describes 1 of the items in detail and 1 item generally in the column High Level of Meta-cognitive Awareness. | The participant describes a strategy used to correct per-formance errors or assumpt-ions they made about a task or strategy used. | The participant generally describes 1 of the 2 aspects in the column High Level of Meta-cognitive Awareness. | The participant does not describe a strategy used to correct per-formance errors or assumpt-ions they made about a task or strategy used. |
| Evaluating Question 8.2 | Regulation | The participant describes their own post hoc analysis of their per-formance and strategy effective-ness. | The participant describes 1 of the items in detail and 1 item generally in the column High Level of Meta-cognitive Awareness | The participant describes their own post hoc analysis of either their per-formance or strategy effect-iveness. | The participant generally describes 1 of the 2 aspects in the column High Level of Meta-cognitive Awareness. | The participant does not describe their own post hoc analysis of either their per-formance or strategy effective-ness. |

The coding took at least two hours for each interview. After reading each section of the transcript, the coder assigned a level from zero (0) to five (5), based on the rubric, to the participant's response. To help prevent miscoding a participant's response, the coders wrote a description based on the coding rubric of areas that the participant did and did not address in their response to each question. The description was used to help identify if a participant's response was correctly assigned a level based on the rubric. During coding, blue highlighters were used to code the regulation of cognition components, and yellow highlighters were used to code the knowledge of cognition components in each transcript. The coders also took notes about each participants' metacognitive awareness level. These notes were used to help each coder write a synopsis about the participant's overall level of metacognitive awareness. The synopses helped to form an apparent view of each participant's metacognitive awareness. Coders also noted participant's responses that were assigned higher levels based on the coding rubric or that responded uniquely.

Using the assigned level from each coder on each interview question, interrater reliability was determined using Cohen's kappa. Cohen's kappa compares the first and second coders' assigned levels for all 18 participants and 10 scored response sections in each interview. The 10 responses sections included participants' responses to Questions 1–7 and Question 8.2, including the three parts of Question 4. Because Question 8.1 is a polar question requiring a yes or no answer, it was not scored by the coders. The interviewer attempted to keep the interviews to 1 hour each, and some participants were unable to answer all of the questions in the allotted time. Thus, the total number of compared items was 173 not 180. The interrater reliability analysis started by checking for miscodes on the transcripts. A miscode was when the coder's assigned level and rationale for that level did not match. The miscodes required the coder to verify and correct their assigned level and rationale. After the miscodes were addressed, the data was entered into SPSS. The Cohen's kappa statistic was then used to analyze the data for interrater reliability. Cohen's kappa indicated that coders had a beyond-chance moderate agreement of 73%, a standard error of .03, and a *p*-value less than .001 in their assigned levels for the participants' responses.

The researcher compiled the coders' codes, notes, and synopses for each participant into their respective group. Each group's synopses consisted of participant's assigned level of awareness on each of the 10 scored response sections from the interview, their responses that characterized the assigned level of awareness, and the coders' notes and synopses. The combination of these items into a group summary helped characterize each group's metacognitive awareness. After each group's summary was complete, their unique metacognitive awareness themes became evident.

## Results

Each of the three groups was found to have its own unique metacognitive awareness phenomenon while performing common teacher practices. The metacognitive phenomena of each group related to the many facets of the group members' respective teacher practices. Regarding the knowledge of cognition component for all questions, only a few participants belonging to Groups 1 and 3 exhibited medium to high levels of metacognitive awareness. Other than these participants, based on the coded transcripts, each group had a slightly different but overall low to medium level of metacognitive awareness in the knowledge of cognition component. The participants in each group frequently discussed their strategies (declarative knowledge) and infrequently discussed how to use the strategies for planning, organizing, monitoring, debugging, and evaluating (procedural knowledge). Occasionally, a participant from Groups 1 and 3 would also discuss why the strategy was used in terms of cognition. Listed below are representative examples of responses to interview Question 3 that were coded for the knowledge of cognition component.

> Right now, the way I do things primarily is I will introduce the topic, and I can either be just showing slides, or a presentation, or I like to if I can show an audio or a video. (Group 3)

> Once they know the information then we will kind of go back and revisit all of that, but in the applicable sense where they're actually going to be performing the task or the skill. (Group 3)

> I sit down with the standard course of study, create a pacing guide, roughly when I'm going to teach what, how much time it's going to take, and then create lessons for each day. (Group 2)

> Typically, what I'll do is I have a PowerPoint slide that I'll throw up. It'll give the day's agenda for the three different sections I've been teaching for that period, and it'll give specific warm ups. (Group 1)

The participants from Groups 1 and 2 focused on a strategy and generally how to use the strategy for planning and organization. The participants' responses from Group 3 included strategies, generally how to use the strategies, and a conditional reason why the strategy was used in terms of cognition. However, when comparing each group's knowledge of cognition, the groups were similar. The groups' transcripts had specific declarative knowledge, usually general procedural knowledge, and often lacked conditional knowledge.

The groups' answers to interview questions based on the regulation of cognition component were more distinguishably different. Group 3 participants were assigned more high and high-to-medium levels, Group 2 participants' levels ranging from medium to low, and Group 1 participants were assigned more high-to-medium and medium levels. In two cases, a participant from Groups 1 and 2 was assigned a level of 0. The 0 level was not on the coding rubric, but both coders in both cases recorded a 0 for those participants' responses. For Questions 1 and 2, coders were looking for the participant to describe planning, goal setting, and allocation of resources. Listed below are representative examples of the regulation of cognition coded responses from interview Question 1 dealing with the planning of instruction.

> With this unit, I used what the state had provided, but I also sought additional resources by using an opportunity to participate in a grant program. This program provided additional resources for the students to use. (Group 3)

> One of the methods I use for planning instruction is called the 5 E's, and its engagement, evaluation, and it's a couple more. (Group 2)

> I draw on my past experience quite a bit. I take a look at what has worked in the past and maybe what hasn't worked so well and then I formulate my lecture, my demonstrations based on that information. (Group 1)

The passage from the Group 3 participant exemplifies how Group 3 members, those who had National Board Certification, often discussed goals or allocation of resources in addition to planning. Group 3 participants would often focus on two items from the rubric and only provide some detail on the third. The Group 1 passage is primarily focused on planning. Group 1 often focused on planning with brief descriptions of either goal setting or allocation of resources. The passage from Group 2 presents a typical answer from participants in Group 2. The combination of incomplete thoughts and little detail on one or two of the rubric items resulted in low assigned levels of metacognitive awareness in the regulation of cognition component for interview Questions 1 and 2.

Questions 3 and 4 were focused on organization. Question 4 was specific to information management, an aspect of organization. In Question 3, coders were looking for the participant to describe implementation of techniques based on an understanding of cognition for organization. Listed below are representative examples of the responses to interview Question 3 that were coded regulation of cognition.

> I would begin with some type of bell ringer. Something to get the students interested in what the topic is for that particular day. (Group 3)

> The overall objective for the lesson [was] to draw the kids interest, to help them make connections to the real world. (Group 3)

> I try to make them aware of where we are today, where we've come from, and hopefully where we're going to go in the future. (Group 2)

> I usually start the class out with some sort of bell ringer, to get them thinking about what it is that we're going to do that day. (Group 1)

> By taking it further and possibly doing a hand-on or application project with it, they're involved. (Group 1)

Group 3 participants, the National Board Certified Teachers, were focused on organization, attending to the cognitive needs of the students to keep them interested. Group 2 participants responded to Question 3 in general terms. The coders had difficulty highlighting any significant responses from Group 2 participants. Due to Group 2's generalized responses to Question 3, their focus seemed to be at the macro level of organization. Group 1 participants' answers to Question 3 were similar to those of Group 3. Occasionally, participants in Group 1 lacked the level of detail shown in Group 3's responses.

Question 4 was split into three scenarios dealing with information management. The first scenario was a general question about interpreting information. The second scenario was specific to information management of nonverbal feedback from students. The third scenario related to information management of new teaching techniques learned from another teacher or from a PD program. In all three scenarios, coders were looking for the participant to describe their cognitive organization, elaboration, summarization, and selective focus on important information. Listed below are representative examples of the responses to interview Question 4 that were coded as regulation of cognition.

> When I receive information, I try to make it relevant to what the situation is as far as my perspective, how I'm going to view it. (Group 3)

> When I receive information, I try to internalize it and fully understand the whole concept without just judgment about the information. (Group 3)

> I try to sometimes put myself in the position of being someone else. (Group 3)

> I get information and sometimes I try to apply it to a project, maybe, the kids are working on. (Group 3)

> Based on your experience or based on your colleagues that you're working with or talking to about it, you can evaluate some of the things as yeah, this is really important piece of information that they need to know, or maybe this particular piece of information is not as critical. (Group 2)

> I try to put it into some type of situation that I'm maybe familiar with. What am I going to do with this information, how am I going to apply it to something I already know or something that I need to know. (Group 1)

Group 3 participants discussed how they organized information, focused on the important information, and either summarized or elaborated on the information that they received. The Group 2 participants mostly talked about one item specifically, and talked about the other aspects of information management more generally or not at all. In the Group 2 passage above, the participant's entire response was specific to focusing on important information, but the participant neglected any real detail about other aspects of information organization. The Group 1 participant's answers were more associated with the elaboration and summarization of information as well as the general management of information.

In Questions 5 and 6, the coders were looking for the monitoring subcomponent of cognitive regulation. Based on the rubric, for a participant's response to be coded as a high level, they needed to describe cognitive assessment of each item in detail. All the participants at least described the monitoring of their students. However, Group 2 often did not describe the cognitive assessment portion. Group 2 discussed general monitoring of their students. Only a few participants described the cognitive assessment of themselves or their methods. Only three participants, one from Group 1 and two from Group 3, described in detail the cognitive assessment of themselves, someone else, and their strategy. Listed below are representative examples of the responses coded regulation of cognition from interview question five, which dealt with the monitoring of cognition during instruction.

> One team was asking the same question and another team was asking the same question, even though they're two different ability levels then that means I missed it somewhere and I need to go back and cover that. (Group 3)

> I will need to change my lesson plan just for that one level and I can remediate with them later or right at that moment. (Group 3)

> I walk my classroom and make sure that the students are doing individual work, or if they are working in learning teams then I know what they are doing. (Group 2)

> I kind of monitor myself because if what I'm saying doesn't make sense to me . . . but I've already said something that after I say that, "wait a minute, what did I just say?" (Group 1)

Question 7 focused on the debugging or adjustment of the teacher's strategy during instruction. The coders were looking for the participant to describe a strategy used to correct performance errors and false assumptions that they had made about the task or strategy being used. None of the participants described the correction of both performance errors and false assumptions with enough detail for both coders to assign a high level of cognitive regulation in the debugging subcomponent. Group 1 provided the most detailed descriptions of their adjustments. Group 1 focused on either errors or assumptions with brief reference to the other aspect of the component. Group 2 received the lowest levels on this component. Both coders wrote that the responses of Group 2 participants were general and sometimes did not seem to provide a direct answer to Question 7. Group 3 answered similarly to Group 1 but lacked the same level of detail. The coders considered Group 3 participants to be more at the medium level of metacognitive awareness for the debugging subcomponent based on the coding rubric. Listed below are representative examples of the regulation of cognition coded responses from interview Question 7.

> I adjust my teaching during the lesson based on how I think the lesson is going . . . whether it's contextual feedback or if it's a spoken-type of feedback or body language. I adjust my lesson just based on what I see if it's working or not. (Group 3)

> I'm trying to make sure that I'm not losing some of them. If I do I try to back up and show a different way to do a particular thing. (Group 2)

> I might backtrack and re-explain something. I might try a different way of explaining something . . . switch people to different groups . . . I might try to find a totally different way to explain something. Some groups, I might do hands-on activities. Other groups, I might show a video or see a PowerPoint, just depending on the dynamics. (Group 1)

Part two of Question 8 was about the participants' cognitive self-evaluation and reflection. The coders were looking for the participants to describe their own post hoc analyses of their performance and strategy effectiveness. Based on the coding of their responses, the groups all had similar metacognitive awareness on question eight; they were all basically assigned at a medium level of metacognitive awareness. Some of the participants provided somewhat more detail than others, resulting in slightly higher levels of awareness being assigned. All participants focused on reflecting either about their performance or their strategy effectiveness. The participants described how, when, and why they tended to reflect. Some of the participants even described what they did based on decisions made during their self-evaluation. Listed below are representative examples of the regulation of cognition coded responses from interview Question 8.2.

> I go back and look at what didn't work, what I need to change. (Group 3)

> The next time I teach this lesson, I'm going to do this. I might leave this part off. That's basically how I plan. (Group 3)

> You have to look back on it, and say, "well, that really didn't go well that way, next time I can try it this way." (Group 2)

> A lot of times driving home I'll think about what I did that day. Did it work? How can I do something a little different? Make it a little better, make it a little more interactive. (Group 2)

> I try to make little notes to myself about what went right, what went wrong, come up with ideas on how I could change things and do thing differently. (Group 1)

## Implications

The first finding from this study indicated that all three groups had similar levels of metacognitive awareness in the knowledge of cognition component. Coders indicated based on the coding rubric that participants typically ranged from medium to low levels of cognitive knowledge. The literature, especially in technology and engineering related PD, has discussed the importance of content and pedagogical knowledge in PD (Bybee & Loucks-Horsley, 2000; Daugherty & Custer, 2012; Mundry, 2007; WestEd, 2000). Content and pedagogical knowledge are often considered foundational characteristics of effective PD (Ball, Thames, & Phelps, 2008; Bybee & Loucks-Horsley, 2000; Mundry, 2007). Underlying both content and pedagogical knowledge is knowledge of cognition; the knowledge of strategies, skills, and abilities that impacts a person's learning ability as well as how and why to use strategies and techniques for increasing performance and accomplishing cognitive tasks. Moreover, research indicates that teachers' knowledge of cognition is linked to their learning ability, pedagogical effectiveness, ability to transfer learning from one context to another, and ability to adapt in a complex educational environment (Bransford et al., 2000; Ertmer & Newby, 1996; Georghiades, 2004; Gourgey, 1998; Hartman, 2001; Kramarski & Michalsky, 2009; Lin et al., 2005; Pucheu, 2008; Wilson & Bai, 2010). This would suggest that teacher PD should focus on metacognitive awareness, including knowledge of cognition, because it will impact many common teacher practices.

The second finding from this study indicated that each group had uniquely different metacognitive awareness in the regulation of cognition component. Overall, Groups 1 and 3 had higher levels in regulation of cognition component, and these two groups successfully completed their PD experience. Self-regulation is considered an aspect within the regulation of cognition component of metacognition (Ertmer & Newby, 1996; Pintrich, Wolters, & Baxter, 2000). Self-regulation has been identified as a factor in a person actively pursuing the learning process (Ertmer & Newby, 1996; Pintrich et al., 2000). Regulation of cognition is a decisive factor in teachers' adaptation ability, which helps in solving problems involving information management and reasoning (Hartman, 2001; Kramarski & Michalsky, 2009; Lin et al., 2005). The complexity that teachers experience each day necessitates their ability to regulate their cognition (Hartman, 2001; Kramarski & Michalsky, 2009; Lin et al., 2005). Regulation of cognition's role in teachers' learning, ability to adapt, and development throughout their career suggests its importance as a focus in PD programs.

## Conclusions

The intent of this research was to understand technology and engineering teachers' level of metacognitive awareness in comparison to their participation and completion of either the T2I2 or the NBPTS program. The findings suggest a connection between teachers' level of metacognitive awareness, successful completion of PD, and ability to manage educational complexity. The results of this study are applicable to future work in improving teacher content and pedagogical knowledge, helping teachers manage educational complexity, helping teachers take an active self-regulated role in PD, and PD effectiveness.

## References

Akturk, A. O., & Sahin, I. (2011). Literature review on metacognition and its measurement. *Procedia Social and Behavioral Sciences*, *15*, 3731–3736. doi:10.1016/j.sbspro.2011.04.364

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407. doi:10.1177/0022487108324554

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press. doi:10.17226/9853

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97–113. doi:10.1177/1468794106058877

Bybee, R. W., & Loucks-Horsley, S. (2000). Advancing technology education: The role of professional development. *The Technology Teacher*, *60*(2), 31–34.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Daugherty, J. L., & Custer, R. L. (2012). Secondary level engineering professional development: content, pedagogy, and challenges. *International Journal of Technology and Design Education*, *22*(1), 51–64. doi:10.1007/s10798-010-9136-2

Denzin, N. L., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

Ernst, J. V., Clark, A. C., DeLuca, V. W., & Bottomley, L. (2013, June). *Professional development system design for grades 6-12 technology, engineering, and design educators*. Paper presented at the 120th ASEE Annual Conference & Exposition, Atlanta, GA. Retrieved from https://peer.asee.org/22373

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional science*, *24*(1), 1–24. doi:10.1007/BF00156001

Georghiades, P. (2004). From the general to the situated: Three decades of metacognition, *International Journal of Science Education*, *26*(3), 365–383. doi:10.1080/0950069032000119401

Gourgey, A. F. (1998). Metacognition in basic skills instruction. *Instructional Science*, *26*(1–2), 81–96. doi:10.1023/A:1003092414893

Guskey, T. R. (2003). What makes professional development effective? *Phi Delta Kappan*, *84*(10), 748–750. doi:10.1177/003172170308401007

Hartman, H. J. (Ed.). (2001). *Metacognition in learning and instruction: Theory, research and practice*. Dordrecht, the Netherlands: Springer. doi:10.1007/978-94-017-2243-8

Kramarski, B., & Michalsky, T. (2009). Investigating preservice teachers' professional growth in self-regulated learning environments. *Journal of Educational Psychology*, *101*(1), 161–175. doi:10.1037/a0013101

Lin, X., Schwartz, D. L., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist*, *40*(4), 245–255. doi:10.1207/s15326985ep4004_6

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Mundry, S. (2007, February). *Professional development in science education: What works?* Paper presented at Professional Development for Engineering and Technology: A National Symposium, Dallas, TX. Retrieved from https://www.researchgate.net/publication/255645159

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, NE: Buros Institute of Mental Measurements.

Prytula, M. P. (2012). Teacher metacognition within the professional learning community. *International Education Studies*, *5*(4), 112–121. doi:10.5539/ies.v5n4p112

Pucheu, P. M. (2008). *An investigation of the relationships between the Scoring Rubrics Inventory and the Metacognitive Awareness Inventory as reported by secondary school core-subject teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3313868)

Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction: Theory, research and practice* (pp. 3–16). Dordrecht, the Netherlands: Kluwer Academic. doi:10.1007/978-94-017-2243-8_1

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Education Psychology*, *19*(4), 460–475. doi:10.1006/ceps.1994.1033

Smylie, M. A., Allensworth, E., Greenberg, R. C., Harris, R., & Luppescu, S. (2001). *Teacher professional development in Chicago: Supporting effective practice*. Chicago, IL: Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/publications/teacher-professional-development-chicago-supporting-effective-practice

WestEd. (2000). *Teachers who learn, kids who achieve: A look at schools with model professional development*. San Francisco, CA: Author.

Wilson, N. S., & Bai, H. (2010). The relationships and impact of teachers' metacognitive knowledge and pedagogical understandings of metacognition. *Metacognition and Learning*, *5*(3), 269–288. doi:10.1007/s11409-010-9062-4

**About the Author**

**Andrew J. Hughes** (andrew.hughes@csusb.edu) is Assistant Professor of Career and Technical Education at California State University San Bernardino.

# Examining the Teaching of Science, and Technology and Engineering Content and Practices: An Instrument Modification Study

***Tyler S. Love, John G. Wells, & Kelly A. Parkes***

**Abstract**

A modified Reformed Teaching Observation Protocol (RTOP) (Piburn & Sawada, 2000) instrument was used to separately examine eight technology and engineering (T&E) educators' teaching of science, and T&E content and practices, as called for by the *Standards for Technological Literacy: Content for the Study of Technology* (International Technology Education Association [ITEA/ITEEA], 2000/2002/2007) and the *Next Generation Science Standards: For States, By States* (NGSS Lead States, 2013). The modified RTOP presented in this article can help provide feedback to teachers at all grade levels concerning their reformed teaching of science content and practices, and T&E content and practices. The instrument achieved acceptable interrater reliability (> 80%) and was tested in a larger study (Love & Wells, 2017). Results revealed a significant difference among participants' teaching of science and T&E content and practices according to a myriad of variables, such as years of teaching experience, years of experience teaching the Foundations of Technology (FoT) curriculum, length of the FoT curriculum training attended, and select preparation experiences. Research implications suggest that the instrument affords equal applicability for examining science educators' teaching of T&E content and practices and informing pre- and in-service teacher preparation efforts by determining key factors that significantly influence educators' teaching of these concepts.

*Keywords:* Technology and engineering education, science education, pedagogical content knowledge (PCK), observation instrument

The main goal of this research was to examine the teaching of science and technology and engineering (T&E) content and practices embedded within Foundations of Technology (FoT), an international T&E education curriculum. To investigate this, an observation instrument was needed to separately rate instructors' teaching of science content and practices and instructors' teaching of T&E content and practices. As a result, a reliable and practical observation instrument was developed to quantify the level at which P–12 educators teach these concepts. This instrument, modified from the widely used Reformed Teaching Observation Protocol (RTOP) (Piburn & Sawada, 2000) and a newly created rubric (Appendix B), helps raters provide more consistent scores for observed teaching practices while providing timely and detailed feedback for instructors to enhance their teaching of science and T&E content and practices.

This research initiates a baseline for effectiveness trials that are necessary to establish the reliability of an instrument for examining the extent to which educators are adequately teaching science and T&E concepts as mandated by both the *Standards for Technological Literacy: Content for the Study of Technology (STL)* (International Technology Education Association [ITEA/ITEEA], 2000/2002/2007) and the *Next Generation Science Standards: For States, By States (NGSS)* (NGSS Lead States, 2013).

Requisite to discussing the instrumentation is establishing a clear definition of the term T&E education to alleviate commonly held misconceptions (International Technology and Engineering Educators Association [ITEEA], 2017). “Technology education . . . provides an opportunity for students to learn about the processes and knowledge related to technology that are needed to solve problems and extend human capabilities” (ITEA/ITEEA, 2000/2002/2007, p. 242). Dugger and Naik (2001) further clarified that

> Technology education is concerned with the broad spectrum of technology, which encompasses, but is not limited to, such areas as: design, making, problem solving, technological systems, resources and materials, criteria and constraints, processes, controls, optimization and trade-offs, invention, and many other human topics dealing with human innovation. (p. 31)

T&E education requires the use of various technologies, materials, and tools to develop engineering solutions, which is distinctly different from educational technology (ITEEA, 2017). Understanding the difference between T&E education and educational technologies is critical for recognizing and accurately recording T&E content and practices ratings when using the instrument presented in this article.

The modified instrument in this study was developed at an important time when science education was experiencing major changes. With the *NGSS* mandating that science educators integrate engineering content and practices within their curricula, it posed some legitimate concerns, specifically how science educators’ would be evaluated on their teaching of engineering content and practices—an area in which they had limited exposure and were not adequately prepared to teach in alignment with the *NGSS* (Nadelson & Farmer, 2012). *NGSS* later developed the Educators Evaluating the Quality of Instructional Products (EQuIP) rubric for measuring and providing constructive criterion-based feedback to better align lessons and units with the *NGSS* (NGSS Lead States, 2016). However, as we present in subsequent sections, the EQuIP rubric cannot be used to adequately examine educators’ separate teaching of science and T&E content and practices. Therefore, a more accurate and reliable instrument was needed.

**Review of Literature**

The philosophical basis for this research is grounded in Shulman's (1987) concept of pedagogical content knowledge (PCK). Since its conception, PCK has been the topic of many notable publications within science education (e.g., Abell, 2008; Baxter & Lederman, 1999; Gess-Newsome, 1999; Hume & Berry, 2011; Loughran, Berry, & Mulhall, 2012). It has also been a controversial research topic, with science education experts questioning its existence (Settlage, 2013) and how accurately it can be examined (Gess-Newsome & Lederman, 1999). Despite these criticisms, research has verified that teachers possess varying degrees of PCK based upon experience and training (Shulman, 2004; Williams & Lockley, 2012). Given that science educators are now expected to teach engineering content and practices, logic dictates that they must have adequate PCK to properly teach engineering concepts within the context of science education. As the science profession moves forward in preparing classroom teachers to address engineering in science education, a reliable instrument will be needed for examining the extent of science educators' PCK in T&E separate from their PCK in science. This research presents an instrument and method to address that need.

Numerous research efforts within science, technology, engineering, and mathematics (STEM) education have examined PCK using a variety of methods, such as classroom observations, interviews, multiple choice questionnaires, assessing student work, and analyzing think aloud teaching tasks (Love, 2013). Despite these findings, there is still no unified agreement among researchers on the best method or methods to assess PCK because of its complexity (Park, Jang, Chen, & Jung, 2011). The authors therefore found it necessary to analyze a multitude of studies from various STEM disciplines in order to construct viable instrumentation for investigating PCK in this study.

Loughran, Mulhall, and Berry (2004) developed the content representation (CoRe) and pedagogical and professional-experience repertoire (PaP-eRs) instruments. The CoRe instrument captures teachers' holistic PCK relative to a specific topic, and the PaP-eRs instrument offers a view into a teaching–learning situation in which the content shapes the pedagogy. They have been widely used in STEM education studies (e.g., Bertram & Loughran, 2012; Hume & Berry, 2011; Rollnick, Mundalamo, & Booth, 2013; Williams, Eames, Hume, & Lockley, 2012; Williams & Lockley, 2012) but are still viewed as imperfect tools due to their topic specificity and time intensive nature. A rubric with three levels of proficiency to rate science teachers' PCK from coded interview responses was created by Lee, Brown, Luft, and Roehrig (2007). Their rubric did not accurately reflect all essential elements of PCK nor did it provide much detail about the specific teaching of content and practices. Park et al. (2011) developed a similar rubric that was used to rate teachers' PCK based on interview responses, lesson plans, and classroom observations. Like Loughran, Berry, and Mulhall (2007), they acknowledged that PCK is topic specific, but it is not efficient to make a rubric for each specific science topic taught. Park et al. admitted that when using a rubric to assess PCK, raters have to make inferences from what the teacher says, believes, and does. They concluded that scoring training is an important yet time intensive component needed when using rubrics to rate teachers' PCK.

The Cognitively Activating (COACTIV) Instruction project (Kunter et al., 2007) developed instrumentation to help distinguish between content knowledge and PCK. The COACTIV instrumentation consisted of 23 paper-and-pencil questions representing scenarios that a mathematics instructor may encounter with their students. PCK ratings from this instrument were limited to participants' paper-and-pencil responses. Furthermore, Gumbo and Williams (2014) investigated the PCK of T&E educators by recording observations every 5 minutes, which created a series of snapshots of PCK elements observed over the duration of a lesson. They triangulated these observation data with both interviews and a content analysis of course textbooks. Although effective, they acknowledged that use of this instrument may not be practical for school systems because it demands a significant amount of time to collect and analyze data.

The most notable instrument considered by the authors was the EQuIP rubric (NGSS Lead States, 2016) because it was aimed toward measuring science educators' alignment and overall quality of lessons and units in relation to the *NGSS*. However, it was determined that the EQuIP rubric was not suitable for this study due to its inability to delineate between the teaching of science content and practices and the teaching of engineering content and practices. Rather, the EQuIP rubric combined science and engineering content and practices by using *and/or* when addressing the teaching of science and engineering concepts (NGSS Lead States, 2016). The EQuIP rubric fails to hold instructors accountable for teaching science and engineering concepts adequately. If using such an instrument to rate observed instruction of both science and engineering concepts, teachers would either rate extremely high or extremely low depending on how the term *and/or* is interpreted by the rater. This is a major reason that the authors chose not to utilize the EQuIP rubric for this study, because it was pertinent to investigate the differences between teaching of engineering and science content and practices. The instrument that the authors found most suitable for this study was the RTOP (Piburn & Sawada, 2000).

**The Reformed Teaching Observation Protocol**

The RTOP (Piburn & Sawada, 2000) could be easily modified to help delineate between the teaching of science and T&E content and practices. It also examined instructors' reformed teaching methods, which aligned well with the *NGSS's* expectations for teaching scientific inquiry and engineering design. The RTOP is grounded in constructivism and was designed as an observational instrument to measure reformed teaching of mathematics and science. Various studies (e.g., Nicholas & Lomas, 2009; Ogletree, 2007; Park et al., 2011) have used it in conjunction with other instruments to measure teachers' PCK. Specifically, Nicholas and Lomas (2009) found that the fourth section of the RTOP was able to assess teachers' understanding of key content and provide valuable insight about their PCK. Piburn and Sawada (2000) and Taylor et al. (2013) found the RTOP to be a reliable and valid instrument aligned with national mathematics and science standards documents (e.g., American Association for the Advancement of Science [AAAS], 1989, 1993; National Council of Teachers of Mathematics [NCTM], 1989, 1991, 1995; National Research Council [NRC], 1996). Because of its alignment with the *NGSS's* recommendation for research-based instructional reform (Taylor et al., 2013), the RTOP was deemed adequate to rate observations in this study.

As previously discussed, although a variety of instruments have been used to evaluate the PCK of science and T&E teachers, none were comprehensive in their evaluation of teacher practices. Given the growing focus in P–12 education on integrating the content and practices from multiple disciplines within a single subject, an instrument robust enough to identify specific instructional areas requiring further pedagogical preparation would be an important mechanism for informing pre- and in-service teacher preparation programs on those experiences identified as necessary for preparing educators to teach cross-disciplinary STEM concepts. Hence, modifications to the RTOP instrument were warranted. The following research questions helped guide the instrument modifications and data collection.

1. How accurately can differences among educators' teaching of science content, science practices, T&E content, and T&E practices each be quantified with a practical and reliable observation instrument?
2. To what extent does instructors' effectiveness in teaching science content, science practices, T&E content, and T&E practices differ according to the type of teacher preparation completed and the amount of teaching experience?

## Methodology

The methodology employed in this study was designed to examine the preparation factors influencing the teaching of science concepts embedded within the FoT curriculum.[1] The authors decided to examine the teaching of T&E educators, as opposed to science educators who were only tasked with teaching new content and practices within the past year. T&E educators have been expected to integrate science and engineering concepts since the release of the *STL* (ITEA/ITEEA, 2000/2002/2007) 16 years ago, making them a more

[1] Details regarding the method employed are those described in previously reported research (see Love & Wells, 2017).

viable population for testing the content and practice items of the instrument. The expectation was that T&E educators' increased experience with teaching these concepts would provide more accurate ratings that could otherwise be impacted by recently being tasked with teaching new content and practices.

**Participants**

This study utilized the same pool of participants as described in Love & Wells (2017). First, an online survey collecting demographic and preparation data was sent to all FoT teachers within 12 county school systems of a mid-Atlantic state. Of the 55 survey respondents, eight were purposefully selected for the classroom observation portion, which utilized the modified RTOP. The purposeful selection ensured that a sample of teachers with varying levels of science and T&E preparation experiences were observed. The demographics and preparation experiences of the observed participants are reported in Table 1.

**Table 1**
*Demographic Data for Observed Participants*

| Part. | Gen. | Eth. | Age | Years of teaching | Years teaching FoT | Years teaching science | T&E education certified | FoT training | Inter-Disciplinary STEM course |
|---|---|---|---|---|---|---|---|---|---|
| T 1 | M | C | 62 | 30 | 2 | 0 | Yes | .5 day | No |
| T 2 | M | C | 47 | 10 | 8 | 1 | Yes | 1 week | Yes |
| T 3 | F | C | 24 | 2 | 2 | 0 | Yes | 1 week | No |
| T 4 | M | C | 47 | 13 | 5 | 2 | Yes | .5 day | Yes |
| T 5 | M | C | 56 | 33 | 4 | 0 | Yes | .5 day | Yes |
| T 6 | M | C | 61 | 28 | 10 | 0 | Yes | 1 week | No |
| T 7 | M | C | 59 | 21 | 6 | 0 | Yes | 1 week | Yes |
| T 8 | M | AA | 25 | 3 | 3 | 0 | No | 1 week | No |

*Notes:* Part. = Participant; Gen. = Gender; Eth. = Ethnicity; T = Teacher; M = Male; F = Female; C = Caucasian; AA = African American.

The observed participants consisted of predominantly White males who were certified to teach T&E education. The mean age of the participants was 48 years, and the average number of years teaching was 18. They had very little if any experience teaching science courses and attended some form of training to learn how to teach the FoT curriculum. About half of the participants reported taking a higher education course that discussed methods for teaching interdisciplinary STEM concepts (e.g., science and engineering). The demographic and preparation data from this research and the full study (Love & Wells, 2017) were consistent with studies examining T&E educators at a national level (Love, 2015; Ernst & Williams, 2015).

**Modifying the RTOP and Developing the Rubric**

Subscale 4 regarding content was modified to accurately examine the content and practices (PCK) of science and T&E concepts. These modifications allowed the data to be consistent with the language and teaching strategies described in the *NGSS*. The RTOP was modified with help from content specialists who had expertise in teacher evaluation, science education, and T&E education. The adaptations began by duplicating Subscale 4 and creating two similar yet separate subscales: one to score teaching of T&E content and practices and one for science content and practices. The words *subject matter* and *concepts* were replaced with the term *content* (Appendix A) to better align with the *NGSS*. This made it easier for raters to distinguish between teachers' content knowledge and pedagogical practices during observations. Additionally, a rubric for each of the Subscale 4 criteria was created to provide clarity regarding what was being observed and to help elicit more consistent ratings (Appendix B). The content specialists suggested modifications to the rubric using language and criteria similar to that provided in the training guide section of the *RTOP Reference Manual* (Piburn & Sawada, 2000). This rubric helped observers determine ratings more easily and consistently while also staying true to the criteria described in the original RTOP.

**Training and Interrater Reliability**

Nicholas and Lomas (2009) established that the use of the RTOP by one trained rater consistently provided a valid assessment of teaching practice in a single classroom observation (about an hour). Due to similar time and funding limitations as experienced by Nicholas and Lomas, one trained rater conducted observations and rated participating teachers in this study. To ensure RTOP rating accuracy and reliability prior to actual use in research observations, content experts were used in establishing an acceptable interrater reliability. Establishing interrater reliability was conducted through two RTOP usage sessions with two content specialists who had expertise in *Integrative STEM Education* (Wells, 2016) teaching practices. All of the raters completed the online RTOP tutorials (Buffalo State University of New York, 2007). Following completion of the online modules, raters were asked to use the modified RTOP to rate two video-recorded FoT lessons from the same units as those observed later during the data collection.

In line with standard protocols for establishing interrater reliability, three rounds were needed to reach consensus with an acceptable level of reliability. The first round had each rater use 10% of the RTOP items to independently analyze the first FoT lesson video, which was followed by a second round using an additional 10% of the RTOP items. At the end of each round, arbitration among raters was conducted to compare, discuss, and justify differences in co-ratings. The same procedures were followed for analyzing the second FoT lesson video, which was followed by a third round using the remaining 80% of the RTOP items. At this point, an acceptable level of interrater agreement, greater than 80% (Howell, 2007), had been achieved (see Table 2), resulting in a viable coding scheme for using the RTOP for scoring a FoT lesson.

**Table 2**
*RTOP Interrater Reliability Percentage Established Among Raters*

| | Observation 1 | | | | Observation 2 | | | |
|---|---|---|---|---|---|---|---|---|
| Round | Rater 1 | Rater 2 | Rater 3 | Total | Rater 1 | Rater 2 | Rater 3 | Total |
| 1 | 80% | 60% | 40% | 60% | 100% | 57% | 57% | 71% |
| 2 | 67% | 50% | 50% | 56% | 100% | 86% | 86% | 90% |
| 3 | - | - | - | - | 86% | 90% | 71% | 83% |

*Note:* Total was the percent agreement among all raters by round.

During the classroom observation, the researcher remained as unobtrusive as possible by taking notes on what they observed regarding the teaching of both science and T&E content and practices. To provide the most accurate RTOP ratings, it was determined that lessons should be recorded for later review. To do this, the teacher was given a lapel microphone, and the audio recording was linked to the researcher's notes using the AudioNote software. This software allowed the researcher to click on any portion of their notes and play the corresponding audio recorded during that part of the observation. Immediately after the completion of each school visit, the researcher used the modified RTOP instrument to rate the overall teaching strategies that they observed during the lesson. To ensure the ratings were as accurate as possible, within 48 hours, the researcher reviewed the lesson audio and their corresponding notes to confirm or adjust the ratings. After all observations were completed, they were again reanalyzed all at once via the audio recordings and notes for consistency across observation ratings.

In modifying the RTOP for use in assessing FoT teaching practices, there are recognized limitations that are worthy of mention. Specifically, observer ratings could only be provided based on criterion specified by the instrument. Although Nicholas and Lomas (2009) found the RTOP acceptable for a single observer to rate one lesson, the ratings in this study only reflect a snapshot of the teacher's full range of instructional practices. When using the RTOP, ratings are dependent upon the rater's knowledge of teaching both science and T&E content and practices. Additionally, the sample of observed teachers consisted of primarily White males, which presents a limitation considering the diversity in the actual population of T&E educators. However, it is noteworthy that the convenience sample used in this study is actually reflective of national T&E

educator characteristics (Love, 2015; Ernst & Williams, 2015). Accounting for recognized limitations, the modified RTOP provided a viable instrument for quantifying teacher levels of science and T&E PCK, which is discussed in the following section.

## Observations and Results

### Relationship Among Observations and Preparation Factors

The sum of all seven RTOP category ratings for each participant ranged from 18 to 118 out of a possible 140 with a mean of 59.4 (42%). Specifically in terms of science content and practices, the mean rating for teaching of science content was 9.6 (48%), and the mean rating for teaching of science practices was 5.8 (29%). Four T&E teachers scored 6 (30%) or lower on teaching of science content, and three received a score of 1 (5%) on the teaching of science practices. Conversely, these teachers scored higher in observed teaching of T&E content and practices, as demonstrated by their mean ratings of 13.6 (68%) for T&E content and 7.6 (38%) for T&E practices. When examining these scores in more detail, there were only three participants who scored 9 (45%) or lower out of a possible 20 for teaching T&E content, but three teachers scored a two (10%) or lower for teaching T&E practices (see Table 3).

**Table 3**
*Participants' Observation Ratings: Scores According to RTOP Categories*

| | Category and score | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Part. | LD&I | SC | T&E C | SP | T&E P | CI | S/TR | Total |
| T 1 | 2 | 6 | 7 | 1 | 1 | 3 | 3 | 23 |
| T 2 | 12 | 9 | 17 | 6 | 12 | 12 | 10 | 78 |
| T 3 | 2 | 3 | 7 | 1 | 1 | 3 | 3 | 20 |
| T 4 | 12 | 17 | 20 | 10 | 16 | 13 | 14 | 102 |
| T 5 | 15 | 19 | 20 | 15 | 16 | 15 | 18 | 118 |
| T 6 | 0 | 3 | 9 | 1 | 2 | 1 | 2 | 18 |
| T 7 | 6 | 14 | 19 | 7 | 6 | 7 | 10 | 69 |
| T 8 | 4 | 6 | 10 | 5 | 7 | 9 | 6 | 47 |
| Mean | 6.6 | 9.6 | 13.6 | 5.8 | 7.6 | 7.9 | 8.3 | 59.4 |

*Notes:* Part. = Participant; T = Teacher; LD&I = lesson design and implementation; SC = science content; T&E C = technology and engineering content; SP = science practices; T&E P = technology and engineering practices; CI = communicative interactions; S/TR = student–teacher relationships. Scores for each category range from 0–20, with higher scores indicating a greater rating.

When analyzing the RTOP scores according to the type of teacher preparation experiences participants had, all three groups received their highest mean ratings in T&E content and their lowest in science practices. Further analysis based on preparation experiences indicated that the participant holding an engineering degree had the highest mean RTOP rating (102), whereas teachers who completed teacher preparation programs in disciplines outside of science and T&E education received the lowest mean score (48; see Table 4).

**Table 4**
*Summary of RTOP Ratings According to Preparation*

| Teacher preparation | Total *n*(μ) | SC(μ) | T&E C(μ) | SP(μ) | T&E P(μ) |
|---|---|---|---|---|---|
| T&E teacher prep | 4(58) | 10.5 | 13 | 6 | 6 |
| Non-T&E teacher prep | 3(48) | 6 | 9 | 2 | 3 |
| Engineering prep | 1(102) | 17 | 20 | 10 | 16 |

*Notes:* SC = science content; T&E C = technology and engineering content; SP = science practices; T&E P = technology and engineering practices. Scores for each category range from 0–20, with higher scores indicating a greater rating.

Similar findings emerged when analyzing the data according to teaching experience categories. Novice teachers (1–5 years) had the lowest mean RTOP rating (33.5), veteran teachers (16 or more years) had the next highest rating (57), and intermediate teachers (6–15 years) had the highest rating (90). Again, all groups recorded their highest mean ratings in T&E content and their lowest in science practices (see Table 5).

**Table 5**
*Summary of RTOP Ratings According to Experience*

| Experience level | Total n(μ) | SC(μ) | T&E C(μ) | SP(μ) | T&E P(μ) |
|---|---|---|---|---|---|
| Novice | 2(33.5) | 4.5 | 8.5 | 3 | 4 |
| Intermediate | 2(90) | 13 | 18.5 | 8 | 14 |
| Veteran | 4(57) | 10.5 | 14 | 6 | 6 |

*Notes:* SC = science content; T&E C = technology and engineering content;

SP = science practices; T&E P = technology and engineering practices. Scores for each category range from 0–20, with higher scores indicating a greater rating.

The breadth of experience levels represented by participating teachers resulted in a range of high and low ratings. Consistent across type of preparation and teaching experience were the low scores in teaching science practices and high scores in teaching T&E content (see Tables 4 and 5). Among all eight participants, only two (Teachers 4 and 5) received a perfect score on any item, which occurred in teaching T&E content. Veteran Teachers 1 and 6 earned two of the lower total cumulative ratings for all seven categories (18 and 23 respectively) among all observed participants, whereas intermediate teachers posted some of the higher summative ratings (102, 78). Furthermore, two of the lowest ratings were received by veteran instructors with over 28 years of teaching experience. It is tempting to make inferences with respect to the pedagogical expectations of veteran teachers, and caution is needed here not to draw inaccurate conclusions. The data in this study were not large enough to draw such conclusions; therefore, further analysis was done in order to identify the types and amounts of preparation experiences that had significant positive correlations with observed teaching of content and practices (Love & Wells, 2017).

## Conclusions and Implications

The overarching goal of this research was to develop an instrument that could separately examine the teaching of science and T&E content and practices. Recognizing the limitations imposed by the number of participants and their demographics, our analysis would suggest that the modified RTOP, together with the criterion-based rubric, has merit for discerning the degree to which the preparation of T&E teachers provides them with sufficient science content and practices to teach the science inherent within lessons such as those for FoT. The data suggest that the instrument could serve as a reliable and feasible observation tool to help school systems better focus their professional development efforts. One of the benefits of the instrument is that it directly measures those instructional strategies observed, as opposed to test question responses, and this research provides some evidence to suggest that the modified RTOP has potential for use with other science or T&E lessons. Further research would be needed to substantiate this potential.

Although observation of T&E participants was limited to a single 1-hour lesson, their limited formal preparation in science content and practices (Love, 2015; Love & Wells, 2017), coupled with low ratings on teaching science concepts in the lessons (Table 3), suggests the need to investigate whether or not similar results would be found for science teachers attempting to teach T&E content and practices inherent within science lessons. Considering that preservice science teacher preparation programs prior to *NGSS* did not address engineering practices, such research might prove useful in efforts to enhance science educators' effectiveness in teaching engineering concepts with the rigor needed for students to make higher level cross-disciplinary connections. Moreover, the ratings found in relation to preparation type (Table 4) and years of teaching experience (Table 5) further emphasize the importance of examining the amount and types of preparation factors that influence instructors' teaching of science and T&E concepts. However, in spite of these findings, further research examining teacher preparation experiences that influence science educators' teaching of T&E concepts is needed to ensure a more holistic preparation of those instructors implementing the *NGSS*. The instrument presented in this article provides a starting point for such research and the potential for highlighting explicit areas in which school systems should focus their professional development efforts.

Our findings have implications for examining the science and T&E PCK of teacher educators who are preparing science and T&E educators. With the recently released *NGSS* and mandated integration of engineering concepts, few would argue that the science educators preparing tomorrow's science teachers may benefit from professional development on how best to integrate engineering concepts within the science curriculum. One readily available solution is collaboration between the science and T&E education programs in preparing both science and T&E teachers with the PCK that they need to teach cross-disciplinary STEM concepts (Wells, 2008).

## References

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, *30*(10), 1405–1416. doi:10.1080/09500690802187041

American Association for the Advancement of Science (AAAS) (1989). *Science for all Americans.* New York, NY: Oxford University Press.

American Association for the Advancement of Science (AAAS) (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 147–161). Dordrecht, the Netherlands: Kluwer Academic. doi:10.1007/0-306-47217-1_6

Bertram, A., & Loughran, J. (2012). Science teachers' views on CoRes and PaP-eRs as a framework for articulating and developing pedagogical content knowledge. *Research in Science Education*, *42*(6), 1027–1047. doi:10.1007/s11165-011-9227-4

Buffalo State University of New York. (2007). Reformed teaching observation protocol. Retrieved from http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/index.htm

Dugger, Jr., W. E., & Naik, N. (2001). Clarifying misconceptions between technology education and educational technology. *The Technology Teacher*, *61*(1), 31–35.

Ernst, J. V., & Williams, T. O. (2015). The “who, what, and how conversation”: Characteristics and responsibilities of current in-service technology and engineering educators. *The Journal of Technology Studies*, *41*(1), 48–56. doi:10.21061/jots.v41i1.a.6

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. . In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 3–17). Dordrecht, the Netherlands: Kluwer Academic. doi:10.1007/0-306-47217-1_1

Gess-Newsome, J., & Lederman, N. G. (Eds.). (1999). *Examining pedagogical content knowledge: The construct and its implications for science education.* Dordrecht, the Netherlands: Kluwer Academic. doi:10.1007/0-306-47217-1

Gumbo, M. T., & Williams, P. J. (2014). Discovering Grade 8 technology teachers’ pedagogical content knowledge in the Tshwane District of Gauteng Province. *International Journal of Educational Sciences*, *6*(3), 479–488.

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thomson Wadsworth.

Hume, A., & Berry, A. (2011). Constructing CoRes—a strategy for building PCK in pre-service science teacher education. *Research in Science Education*, *41*(3), 341–355. doi:10.1007/s11165-010-9168-3

International Technology and Engineering Educators Association (ITEEA). (2017). Technological literacy standards: Technology education vs educational technology. Retrieved from https://www.iteea.org/Activities/2142/Technological_Literacy_Standards/45979/51801.aspx

International Technology Education Association (ITEA/ITEEA). (2000/2002/2007). *Standards for technological literacy: Content for the study of technology* (3rd ed.). Reston, VA: Author.

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Krauss, S., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG Priority Programme* (pp. 32–52). Münster, Germany: Waxmann.

Lee, E., Brown, M. N., Luft, J. A., & Roehrig, G. H. (2007). Assessing beginning secondary science teachers' PCK: Pilot year results. *School Science and Mathematics*, *107*(2), 52–60. doi:10.1111/j.1949-8594.2007.tb17768.x

Loughran, J., Berry, A., & Mulhall, P. (2007). Pedagogical content knowledge: What does it mean to science teachers? In R. Pintó & D. Couso (Eds.), *Contributions from science education research* (pp. 93–105). Dordrecht, the Netherlands: Springer. doi:10.1007/978-1-4020-5032-9_7

Loughran, J., Berry, A., & Mulhall, P. (2012). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam, the Netherlands: Sense. doi:10.1007/978-94-6091-821-6

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, *41*(4), 370–391. doi:10.1002/tea.20007

Love, T. S. (2013). Theoretical underpinnings toward assessing science pedagogical content knowledge (PCK) of technology educators. In J. Williams & D. Gedera (Eds.), *Technology education for the future – A play on sustainability.* Proceedings of the 27th Pupil's Attitude Toward Technology Conference, Christchurch, New Zealand: University of Waikato (pp. 291-296).

Love, T. S. (2015). Examining the demographics and preparation experiences of foundations of technology teachers. *The Journal of Technology Studies, 41*(1), 58-71. doi.org/10.21061/jots.v41i1.a.7

Love, T. S., & Wells, J. G. (2017). Examining correlations between the preparation experiences of U.S. technology and engineering educators and their teaching of science content and practices. *International Journal of Technology and Design Education.* doi: 10.1007/s10798-017-9395-2. Retrieved from http://rdcu.be/o61P

Nadelson, L. S., & Farmer, C. (2012, July 31). Developing standards for teaching engineering. *NSTA WebNews Digest*. Retrieved from http://www.nsta.org/publications/news/story.aspx?id=59528

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM) (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM) (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

National Research Council (NRC) (1996). *National science education standards.* Washington, DC: National Academy Press. doi:10.17226/4962

Nicholas, P., & Lomas, G. (2009). The impact of a university course focusing on PCK and MCK: Does teachers' classroom practice reflect the professional development experience? In *Findings from the New Zealand numeracy* *development projects* (pp. 189–197). Wellington, New Zealand: Learning Media Limited. Retrieved from https://nzmaths.co.nz/sites/default/files/Numeracy/References/Comp09/comp09_lomas_nicholas.pdf

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: National Academies Press. doi:10.17226/18290

NGSS Lead States. (2016). *EQuIP rubric for lessons & units: Science.* Retrieved from http://www.nextgenscience.org/resources/equip-rubric-lessons-units-science

Ogletree, G. L. (2007). *The effect of fifth grade science teachers' pedagogical content knowledge on their decision making and student learning outcomes on the concept of chemical change.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3313734)

Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, *41*(2), 245–260. doi:10.1007/s11165-009-9163-8

Piburn, M., & Sawada, D. (2000). *Reformed Teaching Observation Protocol (RTOP): Reference manual* (ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers, Arizona State University.

Rollnick, M., Mundalamo, F., & Booth, S. (2013). Concept maps as expressions of teachers' meaning-making while beginning to teach semiconductors. *Research in Science Education*, *43*(4), 1435–1454. doi:10.1007/s11165-012-9314-1

Settlage, J. (2013). On acknowledging PCK's shortcomings. *Journal of Science Teacher Education*, *24*(1), 1–12. doi:10.1007/s10972-012-9332-x

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22. doi:10.17763/haer.57.1.j463w79r56455411

Shulman, L. S. (2004). *Teaching as community property: Essays on higher education*. San Francisco, CA: Jossey-Bass.

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2013). *An efficacy trial of research-based curriculum materials with curriculum-based professional development*. Colorado Springs, CO: Biological Sciences Curriculum Study. Retrieved from https://bscs.org/efficacy-trial-research-based-curriculum-materials-and-professional-development

Wells, J. G. (2008, November). *STEM education: The potential of technology education*. Paper presented at the 95th Annual Mississippi Valley Technology Teacher Education Conference, St. Louis, MO. Retrieved from http://www.mississippivalley.org/wp-content/uploads/2015/12/Wells_2008_MississippiValleyConference_STEM-ED_TE-Potential.pdf

Wells, J. G. (2016). PIRPOSAL model of integrative STEM education: Conceptual and pedagogical framework for classroom implementation. *Technology and Engineering Teacher, 75*(6), 12–19.

Williams, J., Eames, C., Hume, A., & Lockley, J. (2012). Promoting pedagogical content knowledge development for early career secondary teachers in science and technology using content representations. *Research in Science & Technological Education*, *30*(3), 327–343. doi:10.1080/02635143.2012.740005

Williams, J., & Lockley, J. (2012). Using CoRes to develop the pedagogical content knowledge (PCK) of early career science and technology teachers. *Journal of Technology Education*, *24*(1), 34–53. doi:10.21061/jte.v24i1.a.3

**About the Authors**

**Tyler S. Love** (tslove@umes.edu) is Associate Professor and Coordinator of Technology & Engineering Education at the University of Maryland Eastern Shore

**John G. Wells** (jgwells@vt.edu) is Associate Professor of Integrative STEM Education at Virginia Tech

**Kelly A. Parkes** (parkes@tc.columbia.edu) is Associate Professor in the Department of Arts and Humanities at Teachers College, Columbia University

## Appendix A

| | Propositional Knowledge: Science Content | |
|---|---|---|
| 6a. | The lesson involved detailed explanations and examples about fundamental science content identified by the curriculum. | 0 1 2 3 4 |
| 7a. | The lesson promoted strongly coherent conceptual understanding of science content. | 0 1 2 3 4 |
| 8a. | The teacher had a solid grasp of the science content presented in the lesson. | 0 1 2 3 4 |
| 9a. | Elements of scientific abstraction (e.g., symbolic representations, theory building) were encouraged when it was important to do so. | 0 1 2 3 4 |
| 10a. | Connections with science content to other content disciplines and/or real world phenomena were explored and valued. | 0 1 2 3 4 |

| | Propositional Knowledge: T&E Content | |
|---|---|---|
| 6b. | The lesson involved detailed explanations and examples about fundamental T&E content identified by the curriculum. | 0 1 2 3 4 |
| 7b. | The lesson promoted strongly coherent conceptual understanding of T&E content. | 0 1 2 3 4 |
| 8b. | The teacher had a solid grasp of T&E content presented in the lesson. | 0 1 2 3 4 |
| 9b. | Elements of T&E abstraction (e.g., symbolic representations, theory building) were encouraged when it was important to do so. | 0 1 2 3 4 |
| 10b. | Connections with T&E content to other content disciplines and/or real world phenomena were explored and valued. | 0 1 2 3 4 |

| | **Procedural Knowledge: Science Practices** | |
|---|---|---|
| 11a. | Students used a variety of means (simulations, drawings, graphs, concrete materials, manipulatives, etc.) to represent science phenomena. | 0 1 2 3 4 |
| 12a. | Students made predictions, estimations and/or hypotheses about key science concepts, and devised means for testing them. | 0 1 2 3 4 |
| 13a. | Students were actively engaged in thought-provoking activity that often involved the critical assessment of science procedures. | 0 1 2 3 4 |
| 14a. | Students were reflective about their science learning. | 0 1 2 3 4 |
| 15a. | Intellectual rigor, constructive criticism, and the challenging of scientific content were valued. | 0 1 2 3 4 |

| | **Procedural Knowledge: T&E Practices** | |
|---|---|---|
| 11b. | Students used a variety of means (models, prototypes, drawings, graphs, concrete materials, manipulatives, etc.) to represent T&E phenomena. | 0 1 2 3 4 |
| 12b. | Students made predictions, estimations and/or hypotheses about key T&E concepts, and devised means for testing them. | 0 1 2 3 4 |
| 13b. | Students were actively engaged in thought-provoking activity that often involved the critical assessment of T&E procedures. | 0 1 2 3 4 |
| 14b. | Students were reflective about their T&E learning. | 0 1 2 3 4 |
| 15b. | Intellectual rigor, constructive criticism, and the challenging of T&E content were valued. | 0 1 2 3 4 |

## Appendix B

| Question # | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *6a & 6b*<br>*Detailed explanations and examples* | Targeted fundamental content was **never** mentioned. | Targeted fundamental content **was rarely** mentioned and was **not a focal point** of the lesson. | The lesson had **very little** focus on targeted fundamental content. | The lesson was focused **to some extent** on targeted fundamental content. | The lesson was focused **entirely** around targeted fundamental content. |
| *7a & 7b*<br>*Coherent conceptual understanding* | Targeted content was **not** interrelated with any other concepts. | Targeted content was **vaguely** interrelated with other concepts to **rarely** increase its meaning. | Targeted content was **vaguely** interrelated with other concepts to **sometimes** increase its meaning. | Targeted content was **vaguely** interrelated with other concepts to **moderately** increase its meaning. | Targeted content was **strongly** interrelated with other concepts to **greatly** increase its meaning. |
| *8a & 8b*<br>*Solid grasp of content* | **No evidence** that the teacher has a solid grasp of the content in the information presented to the class. | Teacher **rarely** illustrates a solid grasp of the content in the information presented to the class. | Teacher **sometimes** illustrates a solid grasp of the content in the information presented to the class. | Teacher **frequently** illustrates a solid grasp of the content in the information presented to the class. | Teacher **regularly** illustrates a solid grasp of the content in the information presented to the class. |
| *9a & 9b*<br>*Elements of abstraction* | Relationships were **never** represented in abstract and/or symbolic ways when it was important to do so. | Relationships were **rarely** represented in abstract and/or symbolic ways when it was important to do so. | Relationships were **sometimes** represented in abstract and/or symbolic ways when it was important to do so. | Relationships were **often** represented in abstract and/or symbolic ways when it was important to do so. | Relationships were **regularly** represented in abstract and/or symbolic ways when it was important to do so. |
| *10a & 10b*<br>*Connections to other disciplines/real world* | Targeted content was **never** connected with content across disciplines or with a real world application example. | Targeted content was **rarely** connected with content across disciplines **or** a real world application example. | Targeted content was **sometimes** connected with content across disciplines or included **an example** of a real world application. | Targeted content was **often** connected with content across disciplines or included **2 examples** of real world applications. | Targeted content was **regularly** connected with content across disciplines and included **more than 2** examples of real world applications. |

| Question # | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *11a & 11b*<br>*Means to represent phenomena* | Students **never** used a variety of practices (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent targeted science phenomena. | Students **incompletely** used a variety of practices (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent targeted science phenomena. | Students **sometimes (once or twice)** used a **complete** practice (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent targeted science phenomena. | Students **often (three)** used a variety of **complete** practices (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent targeted science phenomena. | Students **consistently (four or more)** used a variety of **complete** practices (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent targeted science phenomena. |
| *12a & 12b*<br>*Made Predictions/ Estimations/ Hypotheses* | Students were **not** led to state predictions, estimations, and/or hypotheses associated with the targeted content, and **did not** have to devise ways to test it. | Students were **vaguely** led to state predictions, estimations, and/or hypotheses associated with the targeted content, and **did not** have to devise ways to test them. | Students were **clearly** led to state predictions, estimations, and/or hypotheses associated with the targeted content, and **did not** have to devise ways to test them. | Students were **explicitly** led to state predictions, estimations, and/or hypotheses associated with the targeted content, and **devised** ways to test each. | Students were **explicitly** led to state predictions, estimations, and/or hypotheses associated with the targeted content, and **devised several** ways to test each. |
| *13a & 13b*<br>*Actively engaged in thought-provoking activity and critical assessment* | Students were **never** involved with the investigation, **nor** engaged in thought-provoking activity leading to critical assessment of procedures. | Students **were** involved with the investigation, but **rarely** engaged in thought-provoking activity leading to critical assessment of procedures. | Students **were** involved with the investigation, and **sometimes** engaged in thought-provoking activity leading to critical assessment of procedures. | Students **were** involved with the investigation, and **often** engaged in thought-provoking activity leading to critical assessment of procedures. | Students **were** involved with the investigation, and **regularly** engaged in thought-provoking activity leading to critical assessment of procedures. |

| Question # | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *14a & 14b*<br>*Reflective* | Students were **never** reflective about their learning on the targeted content or concepts in tasks. | Students were **vaguely** reflective about their learning on the targeted content or concepts in tasks with a **vague** prompt and **inappropriate** time allowed. | Students were **minimally** reflective about their learning on the targeted content or concepts in tasks with a **minimal** prompt and **minimal** time allowed. | Students were **clearly** reflective about their learning on the targeted content or concepts in tasks with a **clear** prompt and **adequate** time allowed. | Students were reflective on **multiple occasions** about their learning on the targeted content or concepts in tasks with **clear** prompts and **ample** times. |
| *15a & 15b*<br>*Intellectual rigor/criticism/challenging* | Teacher **never** allows ideas to be presented, challenged, or negotiated by students on the targeted content. | Teacher **rarely** allows ideas to be presented, challenged, or negotiated by the students on the targeted content, but **without** evidence. | Teacher **sometimes** allows **some** ideas to be presented, challenged, or negotiated by the students on the targeted content with **very little** accurate evidence. | Teacher **often** allows a **variety** of ideas to be presented, challenged, or negotiated by the students on the targeted content with **some** accurate evidence. | Teacher **always** allows a **variety** of ideas to be presented, challenged, or negotiated by the students on the targeted content **with** adequate and accurate evidence. |

# The MESA Study

*Cameron D. Denson*

**Abstract**

This article examines the Mathematics, Engineering, Science Achievement (MESA) program and investigates its impact on underrepresented student populations. MESA was started in California during the 1970s to provide pathways to science, technology, engineering, and mathematics careers for underrepresented students and represents an exemplar model of informal learning environments. Using a mixed-method research design of investigation, this exploratory study looks at the relationship between MESA activities and underrepresented students' self-efficacy, interests, and perceptions related to engineering. Evidences for this study includes data from focus-group interviews conducted and results from quantitative data collected using the Engineering, Self-Efficacy, Interests, and Perceptions Survey (ESIPS) instrument. Results from this study suggest that participation in MESA's activities has a positive influence on underrepresented students' self-efficacy, interests, and perceptions related to engineering.

*Keywords:* Informal learning, underrepresented student populations, mixed methods research

Broadening the participation of underrepresented populations in the science, technology, engineering, and mathematics (STEM) fields is a matter of national security and has become an emphasis for national policy (Strayhorn, 2015). Yet, recent studies have provided evidence that efforts to address these shortages in STEM areas have fallen short. In describing patterns of enrollment in STEM majors, studies revealed that "less than 15% of undergraduate degrees in engineering, math, and physical science were earned by African American, Latina/o, or Native American . . . students (NSF, 2013)" (MacPhee, Farro, & Canetto, 2013, p. 348). A recent report revealed a troubling trend for underrepresented student populations entering into STEM majors. Of all the bachelor degrees awarded in 2015, Black students represented only 3% of this population, Hispanic students a slightly better 8%, and female students only 19% of all engineering degrees awarded, numbers that are significantly lower than their representation in the general population (National Science Foundation, National Center for Science and Engineering Statistics, 2015). Continuing efforts have tried to address the proportion of participants in engineering who are women and underrepresented minorities, but the demographics of engineering enrollments continue to fall significantly short of the goals of reflecting the demographics of the overall population (Watson & Froyd, 2007).

The result is a STEM field that remains overwhelmingly White, male, and able-bodied, leaving the available pool of talented women, minorities, and persons with disabilities significantly underrepresented (May & Chubin, 2003). To meet this challenge, it is important to identify factors that may help encourage underrepresented student populations to choose careers in STEM fields.

It can be argued that the lack of engineering understanding and a loss of interest in science and mathematics is contributing to the lack of underrepresented students pursuing engineering careers (Jeffers, Safferman, & Safferman, 2004). To effectively address this problem, educators have sought to create new and innovative pathways for attracting a talent pool to STEM professions that encompasses the diversity evident in the nation's general population (Chubin, May, & Babco, 2005). The National Academy of Engineering's Committee on K–12 Engineering Education released a report that detailed the status of engineering in K–12 education (Katehi, Pearson, & Feder, 2009). In this report, the committee stressed the importance of developing curricula with features that appeal to students from underrepresented groups (Katehi et al., 2009). Scientists, engineers, and scholars should not leave the job of recruiting underrepresented populations to STEM careers solely to K–12 teachers of math and science education (Jeffers et al., 2004). Studies show that formal learning environments have traditionally struggled to effectively introduce STEM content and STEM professions to underserved student populations (Denson, Austin, & Hailey, 2012).

Currently, there is a lack of empirical research on the efficacy of intervention programs to influence underrepresented students (Dyer-Barr, 2014). It is important that comprehensive research studies are employed to help illuminate the practices that are particularly effective in recruiting underrepresented students to STEM careers. One way to address the dearth of literature on best practices for recruiting underrepresented students to STEM careers is to investigate the practices of informal learning environments, particularly those that have been effective in recruiting underrepresented students to STEM careers.

## Informal Learning Environments

Informal learning environments may provide the milieu needed for introducing STEM content to all students, but even more importantly, they may provide a pathway to STEM careers for underrepresented students. It is estimated that students spend 86.7% of their time outside of a classroom (Gerber, Cavallo, & Edmund, 2001). This helps illustrate the importance of informal learning environments and the opportunities that they may provide for the teaching and learning of STEM content. Martin (2004) notes that informal learning environments have been an integral part of education for years and will be critical for transforming the teaching of STEM content in the 21st century. Although the merits of informal learning environments are duly noted, research in this area is sparse and undecided on how these experiences benefit students (Gerber et al., 2001). Beyond anecdotal reporting on informal learning environment experiences, there is little research detailing specific activities and their effect on students. This highlights the need to investigate informal learning environments that effectively teach STEM-based concepts to students. Although there are aspects of the program that are conducted during school hours, MESA formally functions as an afterschool program complementing the work of formal STEM curricula. Inferential studies into the ways that informal learning environments are able to impact underrepresented student populations are of particular importance. The results of investigations that explicate how successful informal learning environments impact underrepresented students will provide insight into how the United States can attract diverse populations to STEM fields.

Chubin, May, and Babco (2005) produced a review of engineering-based informal learning environments and concluded that effective engineering-based informal learning environment "must (1) promote awareness of the engineering profession, (2) provide academic enrichment, (3) have trained and competent instructors, and (4) be supported by the educational system of the student participants" (p. 79). Categorically, informal learning environments fall into three settings: (1) "everyday experiences," (2) "designed settings," and (3) "programmed settings" (Kotys-Schwartz, Besterfield-Sacre, & Shuman, 2011, p. 1). Program settings are characterized by "structures that emulate [or complement] formal school settings—planned curriculum, facilitators or mentors (taking a teaching role), and a group of students who continuously participate in the program [(National Academy of Sciences, 2009)]" (Kotys-Schwartz et al., 2011, p. 2). The learning environment featured in this study, the Mathematics, Engineering, Science Achievement (MESA) program has been identified as an effective informal learning environment and is categorically identified as a programmed setting (Mathematics, Engineering, Science Achievement [MESA], 2017a). Research has shown that students who participate in the MESA program "outperform California public high school students overall in the following categories: completion of advance mathematics and physics courses, grades and performance on college entrance exams [(Building Science and Engineering Talent, 2004)]" (Kotys-Schwartz et al., 2011, p. 2–3). Due to MESA's success as an informal learning environment and its unmatched ability to recruit and retain underrepresented student populations to STEM careers (MESA, 2017a), researchers for this study were interested in examining the aspects of MESA that appealed to their underrepresented student populations.

This article reports on the results of an investigation into the impact of the MESA program on underrepresented student populations. Using a sequential, exploratory, mixed-method research design, this article adds to the literature focused on underrepresented student populations and informal learning environments. This article will first provide the reader with a brief history of the MESA organization followed by the research design framing this study. The article will follow with results from focus-group interviews conducted with MESA participants, which provided eight intriguing themes from the MESA organization and helped informed the design of the Engineering, Self-Efficacy, Interests, and Perceptions Survey (ESIPS) instrument used in this study. Finally, this article will provide results and conclusions from quantitative data collected from over 400 student participants using the ESIPS instrument.

**The MESA Organization**

The first MESA program was founded in 1970 at Oakland Technical High School in Oakland, California with a membership of 25 students. MESA's goal was "to develop academic and leadership skills, raise educational expectations, and instill confidence in California's students" from groups that were "historically underrepresented in engineering, physical science, or other math-based fields in order to increase the number of African American, Latino American and American Indian graduates from a four-year university" (MESA, 2017, para. 1). The MESA effort was supported by the California Public School System, the state Community College System, and the California College System. "There may be other established programs, or programs under development, designed to increase Latino academic achievement in mathematics and science, but none has the longevity, organizational structure, network, and academic rigor as does MESA" (Haro, 2004, pp. 218–219). MESA has been able to achieve these goals despite declining federal and state support.

MESA supports educationally disadvantaged students and minority students in middle schools and high schools by providing pathways to help them succeed in science, mathematics and engineering (Kane, Beals, Valeau, & Johnson, 2004). MESA's goals are to: (1) "increase the number of engineers, scientists, mathematicians, and related professionals at technical and management levels, and (2) serve as a driving force in encouraging minorities and females in achieving success in these fields" (Maryland MESA, 2012). MESA programs are based on a common co-curricular academic enrichment model that includes "academic planning, community service, family involvement, academic enrichment, hands-on engineering activities, career advising, field trips, competitions and workshops"(MESA USA, 2011). MESA programs represent an innovative way of linking a co-curricular learning environment to mathematics, engineering, and science programs within the formal public-school setting to enhance the STEM education of students.

Over the past 40 years, the California MESA program has become a model for MESA-USA, a partnership that now involves MESA programs from nine states that are joined together to support disadvantaged and underrepresented students to improve their academic achievement in math, science, and engineering. MESA-USA members are active in Arizona, California, Colorado, Maryland, New Mexico, Oregon, Utah, Washington, and Pennsylvania. Additional information about the history and status of MESA are available on their website: https://mesa.ucop.edu/about-us/.

MESA "has demonstrated through statistics from their California statewide office that MESA students outperform California public school students overall in the following categories: completion of advanced mathematics and physics courses, grades and performance on college entrance exams [(Building Science and Engineering Talent, 2004)]" (Kotys-Schwartz et al., 2011, pp. 2–3). A review of evaluation reports from after-school science, technology, engineering, and mathematics (STEM) programs, both co-curricular and extracurricular, by the Afterschool Alliance found "that attending high-quality STEM afterschool programs yields STEM-specific benefits that can be organized under three broad categories: improved attitudes toward STEM fields and careers; increased STEM knowledge and skills; and higher likelihood of graduation and pursuing a STEM career" (Afterschool Alliance, 2011, p. 2). Further evidence of the program's impact is that California MESA received the Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring in 2000, an award administered by the National Science Foundation on behalf of the President.

The success of individual MESA programs has been well documented, including a recent external evaluation conducted by John Hopkins University with the Maryland MESA program (Corcoran, Eisinger, Reilly, & Ross, 2014). This study looked at MESA's influence on students interest, but it was limited to one state and included just 77 participants (Corcoran et al., 2014), making the results hard to generalize. There is still a need for empirical research that identifies "appropriate content for informal learning models [or environments] or . . . assess[es] the degree to which these informal experiences impact students" (Kotys-Schwartz et al., 2011, p. 1), particularly across programs in different states. In response to this need, this article will present qualitative and quantitative data to illustrate the ways in which MESA is able to influence students' self-efficacy, interest, and perceptions of engineering. Furthermore, relationships among students' interest, perceptions, and self-efficacy will be explored, and qualitative data will be presented on the benefits of MESA for underrepresented students. This study was designed to examine students' participation and involvement in five activities that are common among MESA programs: field trips, guest lecturers, design competitions, hands-on activities, and career and academic advisement.

## Research Design

### Methodology

This study utilized a mixed-method research design. The purpose of this study was to examine the MESA program and understand features of the program that appeal to underrepresented groups. This work complements the work of Tierney and Farmer (2002), by identifying student-oriented activities within the MESA program that have an influence on underrepresented students' engineering self-efficacy, interest in engineering and perceptions of engineering. In addition, focus-group interviews were conducted in an effort to unpack activity variables within the MESA organization and understand the benefits of the program for underrepresented students. The study was conducted in four MESA-USA states: California, Maryland, Washington, and Utah. This study used qualitative and quantitative measures to answer the research questions. The first research question was addressed in the qualitative portion of the study.

1. What are the benefits of participating in MESA for underrepresented student populations?

The second, third, and fourth research questions were addressed in the quantitative portion of the study.

2. What influences do MESA activities have on students' engineering self-efficacy?
3. What influences do MESA activities have on students' interest in engineering?
4. How are the students' perceptions of engineering influenced by their participation in MESA activities?

This study examined student-oriented activities which can be categorized into five distinct groups: (a) field trips, (b) guest lecturers, (c) design competitions, (d) hands-on activities, and (e) student advisement. These five MESA activities represented the independent variables for this study. The dependent variables for this study included students' self-efficacy, interest, and perceptions related to engineering. The study also examined the influence of the MESA program on outcome factors.

Mixed-method research designs are particularly advantageous when seeking to confirm and cross-validate findings within a single study (Creswell, 2009). This study employed an exploratory design of investigation.

> Exploratory designs begin with a primary qualitative phase, then the findings are validated or otherwise informed by quantitative results. This approach is usually employed to develop a standardized (quantitative) instrument in a relatively unstudied area. The qualitative phase identifies important factors, while the quantitative phase applies them to a larger

> and/or more diverse sample (Creswell and Piano Clark, 2007). (Borrego, Douglas, & Amenlink, 2009, p. 59)

In this study, focus-group interviews were used to help identify important features of MESA, which were applied to a larger sample during the quantitative phase. In this study, the qualitative results helped identify features within the MESA program that appealed to underrepresented student populations. In addition, the focus-group results informed instrument development of the ESIPS instrument.

**Theoretical Framework**

The theoretical framework that guided this study was *social cognitive theory* (SCT), which holds that knowledge acquisition is directly related to observing others within their context of social interactions, experiences, and outside media influences (Bandura, 1988). This framework proposes a relationship between outcome expectations and other behavioral factors such as self-efficacy and interest. SCT is based upon the assumption that human ability is a dynamic attribute and that competence in complex tasks requires both well-developed skills and a strong sense of efficacy to deploy one's resources effectively. *Social cognitive career theory* (SCCT) provides a base for exploring the interaction among personal, environmental, and behavioral influences in career development (Lent, Brown, & Hackett, 1994). This framework is appropriate for this study because of SCT's emphasis on the role that self-efficacy, beliefs, outcome expectations, and goals play in career selection.

**Self-efficacy.** The first construct to be explored in this study is self-efficacy, as defined by Bandura (1977, 1986), which refers to the beliefs about one's ability to execute a given task or behavior in order to attain designated performance. Research has provided evidence that the lack of participation of minorities in STEM careers is due in part to low self-efficacy in science and mathematics. Self-efficacy has been found to be a powerful contributor to the decision to pursue a career in STEM and a major predictor of success in STEM courses (Zeldin, Britner, & Parajes, 2008). Although studies have examined self-efficacy as it related to STEM fields, few have focused specifically on engineering (Lent et al., 1994). However, there is evidence that self-efficacy regarding scientific–technical tasks is predictive of student interest (Brown, Lent, & Larkin, 1989) and academic performance (Hackett, Betz, Casas, & Rocha-Singh, 1992) in STEM fields. Bandura (2006) states that "there is no all-purpose measure of perceived self-efficacy" (p. 307). Sherer et al. (1982) noted that self-efficacy has been primarily thought of as a task-specific belief. Thus, in order to measure engineering self-efficacy, a scale must be created specifically related to the engineering domain. "Self-efficacy scales must be tailored to activity domains and assess the multifaceted ways in which efficacy beliefs operate within the selected activity domain" (Bandura, 2006, p. 310). Although some researchers have attempted to create an accurate measure of "general self-efficacy," arguments still persist about the scales validity as a true measure (Chen, Gully, & Eden, 2001). Sherer et al. (1982) assert that when dealing with specific behaviors, more direct behavioral measures will increase the accuracy of the measurement.

**Interest in engineering.** The second construct in this study is interest. If one seeks to account for the low numbers of underrepresented students in STEM careers (e.g., Babco, 2001), one need only look at the trend of tracking and the placement of minority students in lower academic tracks which has negatively impacted student interest in the sciences (Museus, Palmer, Davis, & Maramba, 2011). Multiple studies describe the importance of interest and its relationship to self-efficacy (e.g., Fouad & Smith, 1996; Hutchinson, Follman, Sumpter, & Bodner, 2006; Wender, 2004). Bandura (1986) suggested that perceived efficacy in people fostered the growth of intrinsic interest, which would remain consistent as long as those interests engaged their personal feelings and offered satisfaction. The decades old trend of placing minorities in lower academic tracks does not foster intrinsic interest and may contribute to shortages of minority representation in fields such as science and mathematics (Babco, 2001; Boyer, 1983). A lack of interest in learning science and engineering may come about if one does not see science or engineering as a viable career option. Researchers in science education have asserted that one reason students from low-income communities are not interested in science is that there is "a 'disconnect' between school and home/community life" (Basu & Calabrese Barton, 2007, p. 467). Currently, research offers few solutions on how to sustain these students' interest. However, Basu and Calabrese Barton (2007) found a "strong connection between a sustained interest in science and science learning environments in which students were able to cultivate relationships with people and in ways that reflected their values of relationships and community" (p. 483).

Carlone and Johnson (2007) found that interest in science or science-related fields had less to do with the subject of science than with the effect that their scientific competence would have on the world. The participants in their study were interested in humanitarian work such as health care—efforts that could change the world in a positive way. Interests, along with self-efficacy and outcome expectations, predict intentions, which in turn lead to choice behaviors including those about careers (Lent et al., 1994; Waller, 2006). Waller (2006) also found that African American students' "math self-efficacy and outcome expectations predicted math interest" (p. 543). Brown et al. (1989) showed that even if there is strong interest in a pursuit, if another option is viewed as more attainable that will be the one to which students will strive. In addition to these findings, Fouad and Smith (1996) found that self-efficacy was a large influence on students' interest. Math and science self-efficacy are included among the factors that impact students' interest in engineering.

**Perceptions of engineering.** The final construct to be explored is students' perceptions about engineering. A student's perception of an occupation along with their self-efficacy in skills associated with that occupation greatly influence the likelihood that the student will pursue the occupation (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001). For some individuals, their perceived efficacy rather than their actual achievement is a key determinant of their perceived occupational self-efficacy and preferred choice of work. In a study of African American females, Carlton Parsons (1997) found that 11 of the 20 interviewees imagined a scientist as an unattractive, nerdy, White male. Their image described the male as having a secondary social life with a perfect family. The image that they described did not represent what most African American students see on a daily basis. In fact, negative attitudes toward engineering and "less positive perceptions of the work engineers do" have been reported as key factors in high attrition rates for aspiring engineering students (Besterfield-Sacre, Atmn, & Shuman, 1997; as cited in Hirsch, Gibbons, Kimmel, Rockland, & Bloom, 2003, p. F2A-7). Changing the public's perception of engineering was a major focus of a study reported in the National Academy of Engineering's (2008) report *Changing the Conversation: Messages for Improving Public Understanding of Engineering.* These factors highlight the need to address the negative perceptions that underrepresented students have of sciences.

## Qualitative Study: Focus-Group Study

### Purpose

Focus-group interviews were conducted with two goals in mind. First, the researchers were interested in understanding the nuances of the MESA organization by unpacking the activity variables of the informal learning environment. Second, findings from the focus-group interviews would inform instrument development for the quantitative phase of the study. The purpose of the focus-group interviews included determining the benefits of participating in MESA's informal learning environment for underrepresented students.

### Methodology

The research team used a focus-group protocol to guide the interview sessions. "Focus groups are used to gather opinions" (Krueger & Casey, 2009, p. 2). They consist of a series of interviews, conducted with five to 10 participants, wherein the researcher attempts to gain a certain perspective from a particular group (Krueger & Casey, 2009). Focus-group interviews are well suited for qualitative studies including grounded theory (Webb & Kevern, 2001). Members of the group are there for member checking, expounding upon participant responses, and adding clarity to group responses.

> Focus group interviews typically have five characteristics or features. These characteristics relate to the ingredients of a focus group: (1) people, who (2)

> possess certain characteristics, (3) provide qualitative data (4) in a focused discussion (5) to help understand the topic of interest. (Krueger & Casey, 2009, p. 6)

In order to ascertain a perspective that was reflective of the MESA program, it was important to establish a "consensus" among group members. For the purpose of this study, researchers felt that focus-group interviews were appropriate. The participants for this study were all members of MESA who provided qualitative data during a focused discussion in an effort to inform the researchers as to the aspects of MESA that were particularly beneficial to their experience.

Focus-group interviews are particularly beneficial when seeking consensus: Interactions among participants enhance data quality because participants provide checks and balances on each other's statements (Patton, 2002). A semi-structured interview technique was employed to collect data. During the focus-group interviews, the interviewer was allowed to digress and probe the students for richer descriptions of activities before returning back to the interview guide to maintain the integrity of the interview process (Krueger & Casey, 2009).

**Participant Selection**

Participants were selected for this study using purposeful sampling. Purposeful sampling is an effective strategy of sampling that allows for the collection of "information-rich" data (Glesne, 2006; Patton, 2002). Advisors for each MESA chapter participating in the study selected participants for the focus groups based on student attendance, achievement, and overall participation in the MESA program. Researchers for this study used a purposeful sampling technique in order to secure participants who could provide insight into the aspects of MESA that were beneficial and understand what students are gaining by their participation. Using a purposeful sample of successful MESA programs, researchers were keen in selecting settings and participants who could help illustrate characteristics of the MESA program that led to student recruitment and retention. It is important to note that this focus-group study was not done in an effort to evaluate the effectiveness of MESA; instead, researchers were investigating the aspects of the program that helped recruit and retain students. Participants were provided with food and refreshments as remuneration for their participation. A total of 28 MESA students from five different schools in the California area participated in the five focus-group interviews. Due to convenience, time constraints, and logistical challenges researchers limited their focus-group populations to schools in California. As an example, over a period of 1 week, researchers rented a car and travelled to six different schools in California to collect the data. The student distribution is as follows: Site 1 provided seven participants, Site 2 provided five participants, Site 3 provided six participants, Site 4 provided five participants, and Site 5 provided five participants. In total, there were 19 females and 11 males for a total of 28 student participants. For the duration of this article, participants involved in the qualitative phase will be referred to as focus-group participants in order to distinguish them from the larger student population.

**Data Gathering**

Two researchers were responsible for conducting focus-group interviews. For each interview, one researcher served as a facilitator, and the other researcher served as a note taker. Because both researchers were well versed in conducting qualitative research, they alternated roles as facilitator and note taker. The focus-group interviews were audio recorded, and notes were taken to ensure that data could be crosschecked with the audio recordings.

The interviews took approximately 1 hour to complete for each focus group. The facilitators posed two open-ended questions.

1. Can you think of one of the best times you have had in MESA?
2. What do you think you are gaining by participating in MESA?

After the first question was asked, the facilitator asked additional probing questions for the purpose of clarification and confirmation. This allowed the participants to answer a multitude of questions with minimal probing from the facilitator. After a number of supplementary questions, the second main question was then posed as a concluding question. Again the process was repeated with the facilitator listening carefully to answers and asking additional or follow-up questions from answers given. The themes formed are the result of four recorded interviews and notes taken from a fifth interview. Technical difficulties prevented transcription of the fifth focus-group recordings.

**Analysis**

To build towards a theory of impact and influence relative to MESA activities and underrepresented students, researchers employed a grounded-theory approach to analyze the focus-group results. Grounded theory is an inductive, comparative, iterative method that is used primarily as a method of data analysis. This strategy is useful when striving to render a conceptual understanding from the data (Charmaz, 2001; Dey, 2004). The grounded-theory approach yields themes that are formed from the grouping of codes according to conceptual categories that reflect commonalities among coded data (Glaser & Strauss, 1967). In this study, researchers examined emergent themes formed from the focus-group participants' responses. This was accomplished by looking at the transcribed recordings and notes that were taken during each interview session. Individual researchers reviewed collected responses and gradually went from coding to categories and eventually theory building, which led to the development of activity components (Harry, Sturges, & Klingner, 2005). Using the research question as a guiding framework, frequency counts were used to determine themes that were recurrent in order to identify emergent themes.

Stemler (2001) contends that frequency counts should be used to determine content of particular interest. In qualitative research, “a summative content analysis involves counting and comparisons, usually of keywords or content, followed by the interpretation of the underlying context” (Hsieh & Shannon, 2005, p. 1277). Emergent themes were verified only once consensus was reached amongst the researchers regarding these themes.

In providing a rich, thick description of the emergent themes, quotes were used to help paint a picture and to provide “good data” (Morse & Richards, 2002). Based on the procedures described, readers can have confidence that the conclusions and themes that emerged from the study are in fact a reflection of the participant’s responses. To establish consensus, researchers met with the principal investigator for the study to discuss emergent themes from the analysis. Themes included in the results were a result of frequency counts to determine recurrence. After recurrent themes were identified independently by each researcher, researchers provided evidence in the form of student quotes to the Principal Investigator (PI) for this study in order to move into an emergent theme. The PI was not involved in collecting or coding the data and was thus able to provide an unbiased decision about whether evidence was sufficient to justify inclusion as an emergent theme. After themes were agreed upon and justified to a third party, headers were developed to easily identify the characteristics of the prescribed theme.

## Results and Discussion

The results of the grounded-theory approach to analyzing the focus-group responses produced eight disparate themes that spoke to the benefit of the MESA program for underrepresented students: (a) informal mentoring, (b) makes learning fun, (c) time management (d) application of math and science, (e) feelings of accomplishment, (f) builds confidence, (g) camaraderie, and (h) exposure to new opportunities.

**Informal mentoring.** A surprising theme emerged from the analysis of data from the focus-group interviews. Participants talked more about their roles as mentors in informal mentoring settings, as opposed to the informal mentoring that they received from MESA teachers and advisors. Participants spoke about mentoring not only their fellow underclassmen but also volunteering with local middle and elementary school.

> Instead of doing the competition and competing, we get to volunteer—we get to help with the um middle schools and um help them make their projects and give them advice. (Group 1)

> We tutor elementary schools too, so there’s a lot of elementary schools around. (Group 3)

Informal mentoring from the MESA advisors and teachers was also mentioned:

> This is a club that like wants people, us, all of us, to succeed in life going into college, succeeding in that, all the advisors, all the teachers, just want to see you achieve, to your best quality. So they're going to help you out and to be the best you can be in succeeding. (Group 2)

**Makes learning fun.** Participants seemed to agree that making learning fun was a key part of MESA's success. They not only spoke of the MESA experiences in reference to learning but also voiced the importance of MESA experiences in changing their perceptions of STEM fields.

> It is actually really fun, you don't fall asleep. Um yeah, you don't fall asleep. It's amazing. (Group 1)

> If we didn't have the fundamental of math—I mean fun, in between there—it would be really boring. (Group 2)

> That's something that MESA shows you at hand. You actually see people—actually see engineers and they're just out there doing their thing, and they're just having fun and they're enjoying it. (Group 3)

**Time management.** Organization and time management emerged as a prevalent theme among the focus groups. When speaking about the benefits of MESA, a participant spoke about the impact of the program stating

> Like MESA, like kind of helped me like I used to be something like get on time, and something like that do some other stuff with MESA and taught me that I should be doing stuff earlier than doing it at the last second. (Group 4)

The competitions also aided in developing time management skills:

> You learn that time is of the essence because we're there working, and then once we get to Saturday academies, or regionals, everything has to be on schedule, or we're running late, you have to turn in project at the certain time, so you're running. (Group 2)

MESA advisors helped participants with the organization necessary for application to college programs:

> (MESA helps) when, there's so many deadlines and applications you have to turn in as a senior for college. (Group 1)

**Application of math and science.** The focus-group participants expressed an understanding of the importance of having opportunities to apply the math and science learned in formal learning environments. One participant explained the integration of the formal and informal learning environments:

> So as I would do MESA, I would get more encouraged and be wait, this is what I was learning in class. So where I would learn something in class, I would use it in MESA, and when I would learn something in MESA I would use it back in my class. You're realizing that this isn't just something you're doing for pointless reasons, but you're doing something with it. (Group 2)

Several participants commented on learning the importance of math and science and also the opportunity for transference of knowledge.

> Well the best experience I've had in MESA has been just overall learning the value and importance of math and science. Because we—we put math and science into like—into all these projects we do. (Group 3)

> And MESA really brings out—really tells you like—it really gives you an experience of what it's used for. Like here we're doing physics in class, and I'm like what am I going to use this for? You know, how does this apply to me? But then once you do the windmill or something like that . . . . (Group 1)

**Feelings of accomplishment.** MESA provided opportunities for participants to achieve outcomes that seemed to be key components in the program. Participants voiced feelings about competing and winning competitions.

> One thing you get is just this immense sense of accomplishment, that you did something and it's not [*sic*] something that you just can't fabricate. (Group 2)

> I get an award, I get this medal on to show that I put that much effort into it. And that's something that MESA does for us. Well to me, it makes me feel accomplished like I actually did something, that I put my work into, and I got something out of it. (Group 3)

> We were doing team math, and when we won first place I was, like yes. So it was a good time for me… (Group 4)

**Builds confidence.** Participants commented on gaining and building confidence from winning competitions and seeing others like them succeed.

> I think I've gained a lot of confidence in myself from MESA, because you do a project, or you give a speech, or you take a math test and you kind of think, I don't know I kind of did okay on that, (Group 4)

> Well now that I've done it since 7th grade, it's been easier. Like I'm more calm I know what I'm doing it's just as I go through it I just learned from 7th grade don't be as nervous, just do the best, and just have a little mental power that, you know, I can do it and I can. (Group 2)

> Like it make a difference for me because like all my dad, um a lot of his friends from college became engineers, so it's kind of cool because you may of his friends that I've met have been male engineers, and like I go to their companies and it's all like guys working there. So it was kind of cool because like not only was she an engineer, but she was like in charge of many projects. And so like it showed like how it didn't matter so anybody could do—be in charge. (Group 1)

**Camaraderie.** Although the participants spoke often about participation and placing in various competitions, these activities were often times not what kept them coming back. One of the prevalent themes that we heard was the camaraderie formed by working on projects and visiting different schools while participating in different events. Below are a couple of examples from different groups expressing such thoughts.

> Like if you're at prelims you just kind of cheer for your school if they win, but when you go to regionals, if your center wins, then you're cheering for them. But it's not like that serious, like at competitions like you want to win of course, that's what you're doing it for but like everyone is kind of you know relaxed, and everything like everyone talks to everyone. It's not like, you know, you don't talk to them because they're your competitor or whatever. It's kind of like oh, you're here too, how did you do this year? (Group 1)

> And when we're doing the trebuchet, we spent countless hours. We would go to our advisors house, stay there from like eight in the morning, and it would be eight at night. And we'd be trying to build it. It would be all the groups and we help each other. (Group 2)

> we interact with other schools, and we're—and well, you get to meet new people when you're doing the same project as they are, and they get to give

> you like what Martha said, and everybody else they get to give you hints on what to do on the project, and then besides that, even though you're competing against them, you make new friends that will help you. (Group 2)

The relationships formed among the groups were paramount to why a number of students stay in MESA:

> Well, I stayed with people I didn't really know during MESA that year that well, because they were juniors and I was just a sophomore. It was kind of fun I stayed with them, got to know each other better, got closer for this year. So that was really fun to like, you know, all hang out there together. (Group 1)

**Exposure to new opportunities.** A number of the participants came from backgrounds that do not afford them opportunities to visit college campuses or to work on projects outside of their classrooms. MESA provided a bridge to those participants, which did not go unnoticed or unappreciated.

> MESA, it gives you so many opportunities, that a person like me, would never have had. Like my parents were always—like they complain about the hours I put in for like my projects, you know, but they're like oh, you should—you should do that because it gives you the opportunity—like gives you an opportunity to like see things that we'll never—that you'll never get to see with us, you know. Because my parents they're not really, like um—they don't know any English so they can't go anywhere, so they never take me anywhere and they're just like yeah, so you should just like do your best. And join things that would allow you to see others things, you know, give you opportunities. And MESA really does that. (Group 1)

> And like one thing is like with engineering and stuff, that it—there's not a club on campus that would allow you to explore that option. There's some for writing and reading, the obvious subjects, but sometimes like engineering is kind of like pushed back because it's math and science, the two most unpopular subjects at a school. And then on top of that, you're asked to do a lot of different projects. And without MESA not a lot of teachers would be willing to have just the fun option of trying this. (Group 2)

> Just this last weekend they took some of us juniors to Chico—Chico University and this took us Sunday night and we slept over there at an apartment that these girls share. And so it's not just the fact that you're there, but you get this feeling like you belong. And it's um—you're part of

> the college. And you get to—you get to experience that even before you go to college. And it was really nice we were just—and then we got to visit the dorms we got to visit around school. We saw students, ex-MESA students from this school and they're really happy they say their classes are super hard, but they're loving it. And it's just really nice. And just MESA we're just like—we're really united when it comes to. (Group 3)

## Quantitative Study

The final phase of this mixed-method investigation included the quantitative portion of the data collection, which included over 700 students responding to the survey. The participants came from 22 high schools in the states of California, Maryland, Utah, and Washington. Responses selected for analyses were from individuals who had participated in MESA during the 2012–2013 and 2013–2014 school years. In addition to logistical challenges, communication proved to be a challenge for many states that had agreed to participate in the study. Due to these challenges, data collection extended to over a 2-year period in order to include a diverse set of underrepresented student groups. Below the researchers describe the selection criteria for the participating schools.

For prospective schools to be considered for this study they had to adhere to the following criteria:

1. The high school must have an established MESA program that provided the five student-oriented activities throughout the school year.
2. Students selected will represent underrepresented populations in engineering.
3. Underrepresentation is defined as having a representation in a particular field that is substantially lower than the representation in the general population.
4. Student participants will have at least one academic year of experience in the MESA high school program.

### Data Collection

After school sites indicated interest in participating in the study, a message describing the study and thanking them for their help was sent to the respective site. The message described the purpose of the study and outlined instructions for the administration of the survey. The advisers were asked to have their students complete the survey during a MESA meeting in a room that provided computer and Internet access. A letter of information (LOI) was included with the request that it be sent home to parents or guardians. Some teachers needed hard copies of the LOI, which were sent to them via mail, and others were willing to print copies from an e-mailed PDF file of the LOI. Advisers were informed that if a parent or guardian did not want their child to participate in the study, the parent or guardian should sign and return the form to the adviser. In such a case, the adviser did not allow the student to take the survey. The MESA programs at collaborating schools received a stipend for participating in the study. Each adviser was given an individual SurveyMonkey® link to enable participants to complete their individual survey responses and submit their responses electronically to the project office. An exception was made for one school because the adviser requested paper copies to overcome the lack of student access to computers. The responses from that particular school were entered into the database by a project staff member.

**Instrument.** Data was collected using the ESIPS instrument. The instrument included three subscales for each construct: the self-efficacy subscale contained 11 items, the perceptions subscale contained 12 items, and the interests subscale contained 14 items. There was also an outcomes section that was developed as a result of the qualitative work, which contained 39 items. The ESIPS instrument is a valid and reliable instrument that is the original work of the authors (Denson, Austin, & Hailey, 2014). Cronbach's alpha was presented as measure of internal consistency or reliability for use with the psychometric measure (Schmitt, 1996). Results of the reliability test provided evidence that the results reported are valid and reliable. Results of the reliability estimates revealed very satisfactory scores for each of the subscales: .93 for the self-efficacy subscale, .85 for the perceptions subscale, .90 for the interest subscale, and .96 for the newly added outcomes subscale. The overall Cronbach Alpha for the survey instrument was .96.

Data collection for the quantitative phase of this study was conducted at two different intervals with over 700 students responding to the survey. Participants came from 22 high schools in the states of California, Maryland, Utah, and Washington. Responses selected for analyses were from those individuals who had participated in MESA during the 2012–2013 school year. The responses from California, Washington, and most of Maryland schools included in this report were obtained during the 2013–2014 school year. The responses from Utah schools and several Maryland schools were obtained at the end of the 2012–2013 school year. Many individuals who started the survey failed to complete the instrument but scrolled quickly through the items and exited the survey without providing any useful data. The database of returns was scanned manually, and partial results from respondents who had not completed at least one subscale of the survey were not included in the analysis. After this screening, the final sample for the study was composed of 484 students.

### Data from Tabulations of Survey Results

This section begins with demographic data to orient the reader to the characteristics of the sample of respondents. The data provides information about the genders of the respondents and their ethnicities.

**Gender.** When asked to indicate their gender, 211 respondents indicated that they were male (43.5%), 239 indicated that they were female (49.4%), and 34 (7%) did not respond to the request for gender identification. The distribution of the genders varied considerably across the years in high school in the sample. Females outnumbered males in freshman and sophomore years, whereas males outnumbered females in junior and senior years.

**Ethnicity.** A total of 184 (38%) respondents indicated that they were Hispanic or Latino/Latina, 58 (12%) indicated that they were White, 26 (5.4%) reported being Black or African American, 117 (24.2%) indicated that they were Asian, 7 (1.4%) reported that they were American Indian or Alaska Native, 13 (2.7%) indicated that they were Native Hawaiian or Pacific Islander, and 79 (16.3%) did not respond to the question.

### Descriptive Statistics

The following tables and figures provide descriptive statistics for MESA participants on the following constructs: self-efficacy, perceptions, and interest. The descriptive statistics are presented in order to provide a snapshot of gender, grade, and ethnic differences for each of the prescribed constructs.

**Gender differences.** Gender differences in the responses to the criterion subscales are consistent across the criterion subscales, with male respondents responding more positively than female respondents on self-efficacy, perception, and interest in engineering. This data is presented in Table 1.

**Table 1**
*Gender Differences*

| | Male | Female |
|---|---|---|
| Self-efficacy | 3.89 | 3.47 |
| Perception | 4.10 | 3.92 |
| Interest | 3.68 | 3.36 |

**Differences according to ethnicity.** There were modest differences in mean responses to each of the criterion subscales among respondents from the respective ethnic groups. Means scores for each ethnic group are provided in Table 2.

**Table 2**
*Mean Subscale Scores by Ethnicity*

| | Hispanic or Latino/ Latina (*n* = 184) | White (*n* = 58) | Black or African American (*n* =26) | Asian (*n* = 117) | American Indian or Alaska Native (*n* = 7) | Native Hawaiian or Pacific Islander (*n* = 13) | Undes-ignated |
|---|---|---|---|---|---|---|---|
| Self-efficacy | 3.57 | 3.76 | 3.83 | 3.75 | 3.34 | 3.31 | 3.73 |
| Perception | 3.97 | 3.98 | 4.05 | 4.06 | 3.95 | 3.86 | 3.93 |
| Interest | 3.48 | 3.35 | 3.67 | 3.67 | 3.26 | 3.61 | 3.64 |

Mean scores on each of the three criterion subscales increased slightly among respondents in the respective ascending grade levels (see Table 3). It is important to keep in mind that the differences did not necessarily occur within a specific school.

**Table 3**
*Mean Subscale Scores by Grade Level*

| | Freshman | Sophomore | Junior | Senior |
|---|---|---|---|---|
| Self-efficacy | 3.48 | 3.68 | 3.74 | 3.79 |
| Perception | 3.84 | 3.98 | 4.03 | 4.15 |
| Interest | 3.36 | 3.55 | 3.56 | 3.59 |

**Outcomes assessment.** As the work of the project evolved, it became increasingly obvious that a variety of other outcomes (not included in the self-efficacy, perceptions, and interest subscales) resulted from MESA participation. However, the identification of those outcomes was not as clearly specified in the literature, and these outcomes were not as readily measured as the three generally recognized constructs—self-efficacy, perception, and interest. Using analysis from focus-group interviews conducted, a fourth subscale was developed for the assessment of these outcomes, which included specific areas of cognitive development, growth in affective dimensions, career choices, plans for career preparation, and continuing personal development. The revised ESIPS instrument was pilot tested with 224 students deriving from the states of Washington, California, and Utah (Denson et al., 2014). The researchers utilized the techniques of principle component analysis in order to reduce the number of items into a "principle component" that would account for the most variance of the observed variables. The Kaiser-Meyer-Oklin (KMO) measure of sampling adequacy was calculated and found to be acceptable (0.951). Bartlett's test of sphericity was significant $\chi 2$ (741) = 6495, $p < 0.001$. The analysis produced four components with eigenvalues greater than one. The Cronbach's Alpha's for the four components are: α1 = 0.928, α2 = 0.932, α3 = 0.897, α4 = 0.894 (Denson et al., 2014). Analysis of the outcomes subscale provided several insights that offer opportunities for further investigations.

***Outcomes of MESA involvement.*** Responses to items on the outcomes subscale were analyzed by gender and ethnicity. The results of the gender analysis are presented in Table 4. Male respondents had a mean score of 3.89 on the outcomes items, and female respondents had a mean score of 3.78.

**Table 4**
*Outcome Means by Gender*

| | Total | Male | Female |
|---|---|---|---|
| Outcome means | 3.82 | 3.89 | 3.78 |

Respondents who identified as Hispanic of Latino/Latina and respondents who identified as Native Hawaiian or other Pacific Islander had the highest mean responses on the outcomes items, 3.88. Black or African American respondents and Asian respondents had mean responses of 3.82, White respondents had mean responses of 3.69, and American Indian or Alaska Native respondents had mean responses of 3.66 on the outcomes items. These results appear in Table 5.

**Table 5**
*Outcomes Means by Ethnicity*

| Ethnic Selection | Mean |
|---|---|
| Hispanic or Latino/Latina | 3.88 |
| White | 3.69 |
| Black or African American | 3.82 |
| Asian | 3.82 |
| American Indian or Alaska Native | 3.66 |
| Native Hawaiian or other Pacific Islander | 3.88 |

**Inferential Statistics**

**Specific areas of MESA activities.** Four categories of MESA activities were studied in more detail: hands-on activities, meeting professionals, student advisement, and field trips. Four facets of each of these activities were explored, each with a statement in this portion of the outcomes instrumentation. Using a 5-point Likert type scale, respondents were asked to indicate whether they *strongly* *disagree* (1 point), *disagree* (2 points), *neither disagree nor agree* (3 points), *agree* (4 points), or *strongly agree* (5 points) with each statement.

**Correlations between participation in MESA activities and means on criteria subscales.** One of the most fundamental questions addressed by this study was the influence of involvement in typical MESA activities and subsequent measures of self-efficacy, perceptions of engineering, and interest in engineering. That question is answered most directly by the significant correlations between the means of the hands-on activities, meeting professionals, student advisement, and field trips sections of the survey and the self-efficacy, perception, and interest subscales. Each category of activities was closely correlated with each of the criterion measures. These correlations are reported in Table 6.

**Table 6**
*Patterns of Responses to Items Assessing Effects of MESA Activities*

| | Hands-on activities | Meeting professionals | Student advisement | Field trips |
|---|---|---|---|---|
| Self-efficacy | 0.46 | 0.35 | 0.26 | 0.18 |
| Perceptions | 0.45 | 0.36 | 0.26 | 0.28 |
| Interest | 0.51 | 0.45 | 0.39 | 0.31 |

*Note.* All probabilities < .05*

The results of the correlation matrix illustrate a positive relationship between MESA's activities and students' self-efficacy, perceptions, and interest of engineering. Closer examination reveals that the most influential activity on the aforementioned constructs is hands-on activities. It should be noted that these comparisons help illustrate the difference in influence for aspects of the MESA program. Results from this analysis do not and should not be seen as predictors of success in other informal learning environments.

**Relationship between categories of MESA activities and criterion measures.** Competitive events are among the optional opportunities available to MESA participants. The survey probed several aspects of the individuals' involvement in MESA-sponsored competitive events. Respondents who indicated that they participated in competitive events had significantly higher mean scores on the self-efficacy, perception, and interest subscales of the survey. These results are included in Table 7.

**Table 7**
*Influence of Participation in MESA Competitions Upon Criterion Measures*

| | Yes ($n$ = 309) | | No ($n$ = 161) | | | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | $t$ | $p$ value |
| Self-efficacy | 3.77 | 0.72 | 3.46 | 0.77 | 4.23 | $p < .05$* |
| Perception | 4.07 | 0.54 | 3.88 | 0.66 | 3.15 | $p < .05$* |
| Interest | 3.60 | 0.75 | 3.31 | 0.76 | 3.94 | $p < .05$* |

*Note*. Respondents' participation was determined by their answer to the following question: "Did you participate in competitions at the club or class level?"

## Conclusion

It was the researchers' intent to provide some insight into the aspects of informal learning environments that may influence students and the type of events and activities that were particularly attractive to underrepresented student populations. Results of the study include the development of the ESIPS instrument, data from focus-group interviews, and quantitative research results. The mixed-method investigation of the MESA program yielded both quantitative and qualitative research findings and provided empirical evidence of MESA's impact on underrepresented student populations. Results from focus-group interviews provided insight into the benefits of the MESA program for underrepresented students. These included: (a) informal mentoring, (b) makes learning fun, (c) time management, (d) application of math and science, (e) feelings of accomplishment, (f) builds confidence, (g) camaraderie, and (h) exposure to new opportunities (Denson, Hailey, Stallworth, & Householder, 2015). The results of the qualitative research informed the development of the ESIPS instrument, which was used for the quantitative research phase (Denson et al., 2014).

The quantitative phase of this mixed-method study included administering the ESIPS instrument to over 700 student respondents. The study was able to use the results of 484 underrepresented students who had participated in MESA. The instrument measured students' self-efficacy, interests, and perceptions related to engineering. Findings from this study appear to corroborate expectations that MESA activities and competitive events make substantial contributions toward the development of MESA participants. Inferential statistics provide evidence that participation in MESA provides positive influences on underrepresented students perceptions, interests, and self-efficacy related to engineering. This study examined the relationship between the aforementioned constructs and MESA's hands-on activities, student advisement, meeting professionals, and field trips. The results revealed that each MESA activity was positively correlated with each of the constructs measured; however, hands-on activities seemed to have the strongest impact on students' self-efficacy, interest, and perceptions related to engineering. The study also revealed that students who were active participants in MESA had higher self-efficacy, interest, and perceptions related to engineering.

The MESA respondents in this study appear to follow the general pattern of positive outcomes attributable to participation in structured extracurricular activities reported by Eccles, Barber, Stone, and Hunt (2003) and by Feldman and Matjasko (2005). The MESA experiences of the respondents were perceived as making contributions to their sense of self-efficacy in engineering, their perceptions of engineering, and their interests in engineering. Participation in MESA activities and MESA competitive events had positive outcomes in many dimensions explored in this study. Active involvement in these competitions and organized activities appears to contribute to the development of self-efficacy in engineering, to more accurate perceptions of engineering as a profession, and to enhancing interest in engineering as a field of study and as a career.

The results of this study have several implications for informal learning environments and formal learning environments alike. The results of this study speak to the role of informal learning environments. Operating outside the constraints of standards-based testing and statewide curriculums, the MESA program seems to be effective in impacting students' knowledge, skills, and affective abilities. Although the findings cannot be characterized as surprising, they do appear to corroborate expectations that MESA activities and competitive events make substantial contributions toward the development of MESA participants. Qualitative research results from this study provide empirical evidence as to mechanisms within the MESA organization that appear to make it effective. This is of particular beneficial for formal learning environments that struggle to introduce STEM-based content to underrepresented student populations. In addition, the ESIPS instrument developed for this study is fully adaptable to other engineering-focused informal learning environments and has proven to be a valid and reliable instrument when used in whole or in part. Finally, qualitative results from this study suggest that MESA's activities have an influence on underrepresented students and that student's benefit greatly from their hands-on experiences. Further research should focus on MESA's ability to influence the teaching and learning of STEM content.

## References

Afterschool Alliance. (2011). *STEM learning in afterschool: An analysis of impacts and outcomes*. Retrieved from http://www.afterschoolalliance.org/STEM-Afterschool-Outcomes.pdf

Babco, E. L. (2001). *Underrepresented minorities in engineering: A progress report*. Washington, DC: American Association for the Advancement of Science. Retreived from http://ehrweb.aaas.org/mge/Reports/Report1/Babco-UnderrepresentedMinoritiesinEngineering.pdf

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1988). Organizational application of social cognitive theory. *Australian Journal of Management*, *13*(2), 275–302. doi:10.1177/031289628801300210

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Parajes & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career tajectories. *Child development*, *72*(1), 187–206. doi:10.1111/1467-8624.00273

Basu, S. J., & Calabrese Barton, A. (2007). Developing a sustained interest in science among urban minority youth. *Journal of Research in Science Teaching*, *44*(3), 466–489. doi:10.1002/tea.20143

Borrego, M., Douglas, E. P., & Amenlink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *Journal of Engineering Education*, *98*(1), 53–66. doi:10.1002/j.2168-9830.2009.tb01005.x

Boyer, E. L. (1983). *High school: A report on secondary education in America*. Harper & Row, Inc., 10 East 53rd Street, New York, NY 10022.

Brown, S. D., Lent, R. W., & Larkin, K. C. (1989). Self-efficacy as a moderator of scholastic aptitude—Academic performance relationships. *Journal of Vocational Behavior*, *35*(1), 64–75. doi:10.1016/0001-8791(89)90048-1

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity an an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218. doi:10.1002/tea.20237

Carlton Parsons, E. (1997). Black high school females' images of the scientist: Expression of culture. *Journal of Research in Science Teaching*, *34*(7), 745–768. doi:10.1002/(SICI)1098-2736(199709)34:7<745::AID-TEA5>3.0.CO;2-M

Charmaz, K. (2001). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 675–694). Thousand Oaks, CA: Sage. doi:10.4135/9781412973588.n39

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, *4*(1), 62–83. doi:10.1177/109442810141004

Chubin, D. E., May, G. S., & Babco, E. L. (2005). Diversifying the engineering workforce. *Journal of Engineering Education*, *94*(1), 73–86. doi:10.1002/j.2168-9830.2005.tb00830.x

Corcoran, R. P., Eisinger, J. M., Reilly, J. M., & Ross, S. M. (2014). *An external evaluation of the Maryland Mathematics Engineering and Science Achievement (MESA) program*. Baltimore, MD: Center for Research and Reform in Education (CRRE), John Hopkins University. Retrieved from https://secwww.jhuapl.edu/MESA/Content/Docs/618-636.pdf

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Denson, C., Austin, C. Y., & Hailey, C. E. (2012, June). *Investigating unique aspects of the MESA program for underrepresented students.* Paper presented at the 2012 ASEE Annual Conference & Exposition, San Antonio, TX. Retrieved from https://peer.asee.org/21613

Denson, C. D., Austin, C., & Hailey, C. (2014, March). *Evolution of the ESIPS instrument*. Paper presented at the annual conference of the International Technology and Engineering Educators Association, Orlando, FL.

Denson, C. D., Hailey, C., Stallworth, C. A., & Householder, D. L. (2015). Benefits of informal learning environments: A focused examination of STEM-based learning environments. *Journal of STEM Education*, *16*(1), 11–15. Retrieved from http://www.jstem.org/index.php?journal=JSTEM&page=article&op=view&path%5B%5D=1893

Dey, I. (2004). Grounded theory. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 81–94). Thousand Oaks, CA: Sage. doi:10.4135/9781848608191.d9

Dyer-Barr, R. (2014). Research to practice: Identifying best practices for STEM intervention programs for URMs. *Quality Approaches in Higher Education*, *5*(1), 19–25. Retrieved from http://asq.org/edu/2014/09/research-to-practice-identifying-best-practices-for-stem-intervention-programs-for-urms.pdf

Eccles, J. S., Barber, B. L., Stone, M., & Hunt, J. (2003). Extracurricular activities and adolescent development. *Journal of Social Issues*, *59*(4), 865–889. doi:10.1046/j.0022-4537.2003.00095.x

Feldman, A. F., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research*, *75*(2), 159–210. doi:10.3102/00346543075002159

Fouad, N. A., & Smith, P. L. (1996). A test of a social cognitive model for middle school students: Math and science. *Journal of Counseling Psychology*, *43*(3), 338–346. doi:10.1037/0022-0167.43.3.338

Gerber, B. L., Cavallo, A. M. L., & Marek, E. A. (2001). Relationships among informal learning environments, teaching procedures and scientific reasoning ability. *International Journal of Science Education*, *23*(5), 535–549. doi:10.1080/09500690116971

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (3rd ed.). New York, NY: Pearson.

Hackett, G., Betz, N. E., Casas, J. M., & Rocha-Singh, I. A. (1992). Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology*, *39*(4), 527–538. doi:10.1037/0022-0167.39.4.527

Haro, R. (2004). Programs and strategies to increase Latino students' educational attainment. *Education and Urban Society*, *36*(2), 205–222. doi:10.1177/0013124503261331

Harry, B., Sturges, K. M., & Klingner, J. K. (2005). Mapping the process: An exemplar of process and challenge in grounded theory analysis. *Educational Research*, *34*(2), 3–13. doi:10.3102/0013189X034002003

Hirsch, L. S., Gibbons, S. J., Kimmel, H., Rockland, R., & Bloom, J. (2003, November). *High school students' attitudes and knowledge about engineering.* Paper presented at the 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, CO. doi:10.1109/FIE.2003.1264689

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. doi:10.1177/1049732305276687

Hutchinson, M. A., Follman, D. K, Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, *95*(1), 39–47. doi:10.1002/j.2168-9830.2006.tb00876.x

Jeffers, A. T., Safferman, A. G., & Safferman, S. I. (2004). Understanding K–12 engineering outreach programs. *Journal of Professional Issues in Engineering Education and Practice*, *130*(2), 95–108. doi:10.1061/(ASCE)1052-3928(2004)130:2(95)

Kane, M. A., Beals, C., Valeau, E. J., & Johnson, M. J. (2004). Fostering success among traditionally underrepresented student groups: Hartnell College's approach to implementation of the Math, Engineering, and Science Achievement (Mesa) Program. *Community College Journal of Research and Practice*, *28*(1), 17–26. doi:10.1080/10668920490251944

Katehi, L., Pearson, G., & Feder, M. (Eds.). (2009). *Engineering in K–12 education: Understanding the status and improving the prospects.* Washington, DC: National Academies Press. doi:10.17226/12635

Kotys-Schwartz, D., Besterfield-Sacre, M., & Shuman, L. (2011, October). *Informal learning in engineering education: Where we are—Where we need to go.* Paper presented at the 41st ASEE/IEEE Frontiers in Education Conference, Rapid City, SD. doi:10.1109/FIE.2011.6142836

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). Thousand Oaks, CA: Sage.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*(1), 79–122. doi:10.1006/jvbe.1994.1027

MacPhee, D., Farro, S., & Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: Gender, ethnic, and social class patterns. *Analyses of Social Issues and Public Policy*, *13*(1), 347–369. doi:10.1111/asap.12033

Martin, L. M. W. (2004). An emerging research framework for studying informal learning and schools. *Science Education*, *88*(S1), S71–S82. doi:10.1002/sce.20020

May, G. S., & Chubin, D. E. (2003). A retrospective on undergraduate engineering success for underrepresented minority students. *Journal of Engineering Education*, *92*(1), 27–39. doi:10.1002/j.2168-9830.2003.tb00735.x

Maryland MESA (Mathematics, Engineering, Science Achievement). (2012). Engineering competitions for K-12 students. Retrieved from http://competitions.umd.edu/content/maryland-mesa-mathematics-engineering-science-achievement

Mathematics, Engineering, Science Achievement. (2017). Milestones in MESA history: 1970. Retrieved from https://mesa.ucop.edu/about-us/mesa-history/

Mathematics, Engineering, Science Achievement. (2017). About us. Retrieved from https://mesa.ucop.edu/about-us/

MESA USA. (2011). Mathematics, Engineering, Science Achievement. Retrieved from https://archive.is/0zxA

Morse, J. M., & Richards, L. (2002). *Readme first for a user's guide to qualitative methods*. Thousand Oaks, CA: Sage.

Museus, S.D., Palmer, R.T., Davis, R.J., & Maramba, D.C. (2011). Racial and ethnic minority students' success in STEM education [Special issu]. *Ashe Higher Education Report, 36* (6). doi: 10.1002/aehe.3606

National Academy of Engineering. (2008). *Changing the conversation: Messages for improving public understanding of engineering*. Washington, DC: National Academies Press. doi:10.17226/12187

National Science Foundation, National Center for Science and Engineering Statistics. (2015). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. Arlington, VA. Retrieved from https://www.nsf.gov/statistics/wmpd/archives/wmpd-2015.zip

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Pyschological Assessment*, *8*(4), 350–353. doi:10.1037/1040-3590.8.4.350

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Pyschological Reports*, *51*(2), 663–671. doi:10.2466/pr0.1982.51.2.663

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, *7*. Retrieved from http://pareonline.net/getvn.asp?v=7&n=17

Strayhorn, T. L. (2015). Factors influencing Black males' preparation for college and success in STEM majors: A mixed methods study. *The Western Journal of Black Studies*, *39*(1), 45–63.

Tierney, P., & Farmer, S., M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal*, *45*(6), 1137–1148. doi:10.2307/3069429

Waller, B. (2006). Math interest and choice intentions of non-traditional African-American college students. *Journal of Vocational Behavior*, *68*(3), 538–547. doi:10.1016/j.jvb.2005.12.002

Watson, K., & Froyd, J. (2007). Diversifying the U.S. engineering workforce: A new model. *Journal of Engineering Education*, *96*(1), 19–32. doi:10.1002/j.2168-9830.2007.tb00912.x

Webb, C., & Kevern, J. (2001). Focus groups as a research method: A critique of some aspects of their use in nursing research. *Journal of Advanced Nursing*, *33*(6), 798–805. doi:10.1046/j.1365-2648.2001.01720.x

Wender, I. (2004). Relation of technology, science, self-concept, interest, and gender. *Journal of Technology Studies*, *30*(3), 43–51. doi:10.21061/jots.v30i3.a.7

Zeldin, A. L., Britner, S. L., & Parajes, F. (2008). A comparative study of the self-efficacy beliefs of successfull men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching*, *45*(9), 1036–1058. doi:10.1002/tea.20195

## About the Author

**Cameron D. Denson** (cddenson@ncsu.edu) is Assistant Professor of Technology, Engineering, and Design Education at North Carolina State University.

# Impact of Instructor Teaching Style and Content Course on Mathematics Anxiety of Preservice Teachers

*Suriza Van der Sandt & Steve O'Brien*

**Abstract**

Integrative-STEM methodologies entail integrating multiple disciplines with active design-centric teaching and learning methods. If math anxiety is prevalent, for teachers or students, then both the level of integration and design thinking may be limited. This quantitative study of 160 preservice teachers investigated how math anxiety was impacted by (a) a required math content course, (b) instructor teaching style, and (c) academic and disciplinary major. Significance analyses included *t*-tests, nonparametric tests, and effect sizes. Two teaching styles were compared: a direct teaching style and a more active, problem-based teaching style. The problem-based teaching style was shown to have substantial beneficial impact on math anxiety.

*Keywords:* STEM education, mathematics anxiety, teaching styles

Previous works have discussed the acronym STEM (science, technology, engineering, and mathematics) and K–12 STEM education in general (Sanders, 2009). Sanders (2009) and Virginia Tech (2017) faculty have discussed and defined the term *integrative-STEM education*. Additionally, the National Academy of Engineering and the National Research Council (Honey, Pearson, & Schweingruber, 2009) produced a detailed report describing many aspects of integrative-STEM (I-STEM) methods. This report addressed definitions of I-STEM, reviewed research related to I-STEM education, and discussed practice and implementation of I-STEM. Researchers have also created frameworks to guide I-STEM teaching. Wells (2016) proposed the PIRPOSAL model (which stands for Problem Identification, Ideation, Research, Potential Solutions, Optimization, Solution Evaluation, Alterations, and Learned Outcomes) that has clear ties to problem-based learning (PBL) via the central importance of questioning. Several K–12 school districts have chosen to add an *A* (arts), engaging via STEAM education, peaking the interest of art educators (Liao, 2016).

In the authors' view, if the key attributes of I-STEM teaching and learning could be compressed into two concepts, those would be (a) integrative and (b) include substantial design-centric problem- or project-based learning. These two aspects are not independent but are linked because the design-centric theme (the *T* and *E* components) provides rich contexts for the integration of STEM and non-STEM content areas. The problem- or project-based teaching and learning methods in I-STEM activities are design-centric with teachers guiding a student-centered environment where students, typically working in small groups, are designing solutions to problems, resulting in artifacts representing the solution (a physical artifact or modified process). There are a variety of items that can compromise the quality of I-STEM teaching. For example, questioning techniques are clearly important and have a central place in Wells' (2016) PIRPOSAL model. Another potential factor is mathematics anxiety. The literature indicates that high math anxiety can have several detrimental impacts in the classroom. Each of the two fundamental aspects of I-STEM methods previously listed could be detrimentally impacted. For example, if teachers have high math anxiety, then I-STEM activities may be limited in both the amount and quality of integrated math or may not encourage quantitative design decisions. Additionally, students with high math anxiety may also purposefully shy away from quantitative-based processes.

The authors could find no reported work on the impact of PBL teaching styles on math anxiety of in- or pre-service teachers. This is potentially of fundamental importance to PBL-centric I-STEM classes. For example, if PBL methods can beneficially impact math anxiety of preservice teachers, then perhaps PBL-centric I-STEM methods will also have a beneficial impact on K–12 students and teachers. In this work, a quantitative measure of math anxiety is completed for early preservice teachers before and after a required math (for educators) content course. Independent variables studied are: (a) a required content math course, (b) teaching style (active or PBL vs. direct), and (c) academic and disciplinary majors.

## Literature Review

### Math Anxiety

Mathematics anxiety can be defined as an intense feeling of anxiety about one's ability to understand and do math, a specific event such as a math test, or certain situations involving math. According to Brown, Westenskow, and Moyer-Packenham (2011), math anxiety reflects how an individual views his or her own ability to interact with mathematics.

More broadly, mathematics anxiety can be defined as the stress of learning and participating in the mathematics classroom or in situations that require mathematics (Richardson & Suinn, 1972) or as a fearful avoidance of mathematical situations (Wadlington & Wadlington, 2008). Math anxiety is the result of a student's previous negative or embarrassing experiences with math or a math teacher. Math anxiety is not a learning disability, but it does interfere with an individual's ability to learn math (Wadlington & Wadlington, 2008) and inhibits students' ability to understand and participate in mathematics. Isiksal, Curran, Koc, and Askun (2009) also found a significant negative correlation between math anxiety and self-concept scores. These experiences can leave students with the belief that they are deficient in math. Ashcraft (2002) believes that students with math anxiety will avoid situations requiring math, which could "result in less competency, exposure and math practice, leaving students more anxious" (p. 173). Brady and Bowd (2005) found that nearly 40% of the education students in their study reported math as their least favorite subject.

Math anxiety can develop early in elementary school (Harper & Daane, 1998). Jackson and Leffingwell (1999) reported that some students had their first negative experiences as early as third or fourth grade. Geist (2010) believes that

> Instead of helping children develop fluency at computation and become more efficient at problem solving, these policies [current educational policies] have produced students that rely more on rote memorization and have increased the level of anxiety in young children by making mathematics a high-risk activity. This tends to produce more adults with 'math anxiety' and discouraged children who understand the concept but work a little slower. (p. 25)

Finlayson (2014) believes that teacher behavior is a prime factor contributing to math anxiety.

**Math anxiety of preservice teachers and impact on teaching and learning.** A significantly larger percent of preservice teachers report experiencing higher levels of math anxiety than other undergraduate university students (Harper & Daane, 1998; Hembree, 1990). Frank (1990) found that many future teachers shared many of the same math beliefs held by students enrolled in math anxiety clinics. There is a particular concern in the case of elementary school teachers because a disproportionately large percentage of them experience significant levels of mathematics anxiety (Buhlman & Young, 1982; Trujillo & Hadfield, 1999). Kelly and Tomhave (1985) found that prospective elementary school teachers scored higher on anxiety rating scales than any other group in the large group of college freshmen they tested. Based on this research, it is not surprising that a considerable proportion of students entering preservice teacher training have negative beliefs and attitudes about mathematics (Uusimaki & Nason, 2004). Brown et al. (2011) also believe that teachers who do not enjoy math and who have negative feelings and less ability in mathematics would have difficulty teaching math or teaching math well.

The vast majority of elementary education majors are female and exhibit the highest level of math anxiety of any major (Hembree, 1990). Beilock, Gunderson, Ramirez, and Levine (2010) found that "teachers with high math anxiety seem to be specifically affecting girls' math achievement—and doing so by influencing girls' gender-related beliefs about who is good at math" (p. 1862). Geist (2010) found that "girls tend to feel less confident about their answers on tests and often express doubt about their performance" in math, and over time, girls' "assessment of their enjoyment of mathematics falls much more drastically than" boys' (p. 26). Moreover, studies have shown that teachers with high levels of math anxiety tend to transfer this anxiety to their students (Finlayson, 2014; Vinson, 2001). Some researchers found that such teachers are viewed as unsympathetic (Cornell, 1999) and insensitive (Jackson & Leffingwell, 1999), and Brady and Bowd (2005) found that such teachers were viewed as hostile and uncaring by their students. Furthermore, these students had memories of struggling with particular concepts and experiencing embarrassment in front of peers. Jackson and Leffingwell (1999) report that girls were ridiculed more often than boys and received less assistance from such teachers. Swetman, Munday, and Windham (1993) indicate that teachers with high measures of math anxiety spend less time planning mathematics lessons and use math instruction time for nonmath-related activities.

Additionally, Teague and Austin-Martin (1981) found that a teachers' attitude toward mathematics may affect not only the students' values and attitudes toward mathematics but also that these attitudes may affect the effectiveness of the teaching itself. Brown et al. (2011) established in their study that nearly 21% of the preservice teachers with anxiety about mathematics had negative mathematics teaching experiences with students in their field-based practicum.

Much of the research focusing on math anxiety and preservice-teacher training links math anxiety to teacher efficacy. Swars, Daane, and Giesen (2006) as well as Bursal and Paznokas (2006) found negative correlations between math anxiety and math efficacy beliefs, whereas Gresham (2008) associates low math anxiety with high levels of math efficacy. Math anxiety is also negatively correlated with confidence to teach math (Brady & Bowd, 2005). A commonality among these studies is that the participants were all in a methodology class (e.g., Brown, Westenskow, & Moyer-Packenham, 2011; Finlayson, 2014; Gresham, 2007) and near the end of their preservice training (Brady & Bowd, 2005; Isiksal, Curran, Koc, & Askun, 2009). We are in agreement with Brown et al. (2011) and Finlayson (2014) who acknowledge a weak mathematical background as a factor contributing to math anxiety. Therefore, we decided to investigate the level of math anxiety as they enter their training as teachers. Additionally, we decided to study the impact of a required math content course taken early in their program.

Another limitation in the current literature is the near exclusive focus on math anxiety among elementary major preservice teachers (Brown et al., 2011) with no consideration of other education majors (e.g., early childhood, special education, or deaf and hard of hearing majors) or academic major (e.g., math major). Zientek, Thompson. and Yetkiner (2010) believe that "it may be of value to investigate whether preservice teachers' mathematics anxiety levels are most associated with areas of certification preparation (e.g., early childhood, K–8) or by the courses in which the teachers are enrolled" (p. 430). This investigation includes these other education major populations with additional variables of disciplinary major and instructor's teaching style. By reference, Finlayson's (2014) study, 40% of the study participants identified "teaching style" as a cause for their math anxiety.

### Teaching Style

**Problem-based learning.** Based on the work of Barrows (Barrows, 2002; Hmelo-Silver & Barrows, 2006), Walker and Leary (2009) define a PBL teaching style in which (a) "*ill-structured problems* are presented" (p. 13), (b) "a *student centered* approach in which students determine what they need to learn" is used, (c) "*teachers act as a facilitators* or tutors in the learning process," and (d) "*authenticity* forms the basis" in the selection of "problems [that] are inherently cross-disciplinary" (p. 14). In the work of Barrows (2002) and Hmelo-Silver and Barrows (2006), the importance of group work is included as a fifth element.

Researchers have found that PBL or PBL-like activities have had substantial positive impacts on student learning. PBL was first widely reported in the field of medical education. Meta-analysis of PBL have been reported both in the medical field (Strobel & van Barneveld, 2009) and outside the medical field (Walker & Leary, 2009). Strobel and van Barneveld (2009) found meaningful effect sizes for (a) "knowledge assessment," (b) "performance or skill-oriented" assessment, and (c) "non-performance or skill-oriented" attributes, including "satisfaction" and "successful assignment of first choice of [medical] residency" positions (p. 52). The meta-analysis of Walker and Leary (2009) included the addition of nonmedical field studies and variables of problem types, disciplines, and assessment levels. This meta-analysis described a large number of factors with statistical validity, too many to review here; however, it is interesting to note that the problem type "design problem" had the largest effect size (0.74), which may bode well for I-STEM methods.

**Direct teaching.** Mercer, Lane, Jordan, Allsop, and Eisele (1996) define *explicit* or *direct* instruction as "instruction in which the teacher serves as the [primary] provider of knowledge" and explanations, presenting "skills and concepts . . . in a clear and direct fashion that promotes student mastery" (p. 227). Additionally, Burton (1998) observed that college-level engineering lectures generally take a "'teaching is telling'" approach (p. 158).

## Research Questions

We designed this study to answer the following questions:

1. What is the level of math anxiety with which prospective grade school teachers enter their teacher-training program?
2. What effect does a mathematics content course have on the level of math anxiety experienced by prospective teachers?
3. What effect does the lecturer's teaching style have on the level of math

anxiety?

4. Do different education or disciplinary majors have substantially different math anxiety?

## Methodology

### Population

The population consisted of 160 preservice teachers. Participants were primarily freshmen at a public liberal arts college situated on the East Coast. The mean quantitative SAT scores for education majors at the institution has varied between 600 and 630 over the past 8 years. The population in this study was made up of the following education majors: elementary (ELEM, $n$ = 79), early childhood (EACH, $n$ = 36), deaf-and-hard of hearing (DEAF, $n$ = 23), and special education (SPED, $n$ = 22). EACH students would be certified to teach K-3, ELEM students would be certified to teach K-6, and DEAF and SPED students would be certified to teach K–12. A second, disciplinary major is required for all education majors. The 160 preservice teachers in this study also spanned the following disciplinary majors: art (AR, $n$ = 2), English (ENG, $n$ = 40), history (HIS, $n$ = 16), math (MATH, $n$ = 2), music (MU, $n$ = 3), psychology (PSY, $n$ = 44), sociology (SO, $n$ = 12), Spanish (SPA, $n$ = 6), women and gender studies (WG, $n$ = 11), and integrative STEM (I-STEM, $n$ = 18). Four students were double disciplinary majors (for example, WG and HIS or WG and SPA), one student was a business major, and another student was an international studies major. MATH majors are certified to teach math for K–12, and approximately 90% of I-STEM majors complete the state-required coursework for middle school endorsements for both mathematics and science. Additionally, approximately 50% of I-STEM majors complete coursework for K–12 endorsement for technology and engineering (T&E) education.

### Data Collection and Math Anxiety Instrument

Data were collected for two sequential semesters from students attending a compulsory math content course for elementary school teachers. The content courses were taught by only two instructors: 93 students attended Instructor A's class, and 67 students attended Instructor B's class. At the beginning and end of their course, participants were asked to voluntarily complete the Revised-Mathematics Anxiety Survey (R-MANX), created by Bursal and Paznokas (2006), enabling paired statistical analyses. Only paired data were utilized, measuring predominately the impact of the course (and teaching style). The R-MANX instrument contains 30 items to which students respond on a Likert scale from 1 (*no anxiety*) to 5 (*high anxiety*). Possible scores range from 30–150 with higher scores indicating higher math anxiety. The survey asked the student to define their level of math anxiety when dealing with daily situations and their own coursework. Cronbach's alpha for the R-MANX was calculated as 0.90 (Bursal & Paznokas, 2006).

**Math Content Course Overview**

The compulsory content course is designed for future teachers and is taught by mathematics education faculty. The course explores elementary school mathematics from an advanced viewpoint. Preservice teachers study patterns, numeration, mathematical systems, real numbers, and number theory. Students are required to reason mathematically, solve problems, and communicate mathematics effectively at different levels of formality, using a variety of representations of mathematical concepts and procedures. The two instructors used the same textbook and covered the same chapters. The format for both classes was in-person instruction for approximately 14 weeks. Classes were held twice per week, and each session was 80 minutes long.

**Teaching Style Determination**

One of the researchers visited each of the instructors' classrooms on several occasions during the year to collect data (at least three times per semester). During classroom observations, the researcher took observation notes about the classroom discourse and teaching style demonstrated by each instructor and also collected copies of the syllabi and assessments. Hence, this study was an ex post facto study design.

Instructor A, with 8 years college-level teaching experience, used a variety of physical materials and models (e.g., Cuisenaire rods, pattern blocks, tangrams, and different base blocks). Students were encouraged, through activities based on exploration, to demonstrate a willingness and ability to solve various types of mathematical problems using appropriate strategies. Students were required to explain their answers, reasoning, and problem-solving methods in class, on homework, and on assessments. Students often left their seats to collect manipulatives, work with other students on solving problems, and make brief presentations based on their explorations. Students were often encouraged to work in pairs or groups to explore, discover, and present solutions. The majority of class sessions contained one of more of the five key PBL elements presented earlier in the literature review section. It was evident that Instructor A followed a more problem-based, inquiry-driven teaching style and is referred to as a problem-based teaching style (PBT).

Instructor B, with more than 30 years of college level teaching experience, followed nearly the same procedure at each observation. The instructor presented some example (or examples) on the board, showed the steps in solving the problem (or problems), allowed time for questions, and then assigned homework (problems similar to the example or examples presented in class) to be completed in class and at home. The classroom discourse was instructor driven and blackboard and textbook dominated; no use of manipulatives was observed. Students stayed in their seats, took notes, and worked individually on assigned problems. The teacher set the pace of the discourse, with a clear focus on computation and skill in both teaching and in assessment events. From the observation notes, it is evident that Instructor B followed a primarily direct teaching style, with very little evidence of PBL. In this study, Teacher B is referred to as having a direct teaching style (Direct T).

**Results**

Analyses were completed only for teacher candidates that successfully completed surveys both before and after the math content course, primarily assessing the impact of the course (including teaching style). To test for normality, the Kolmogorov-Smirnov tests were utilized (with $p > 0.05$). If paired data are normally distributed, paired *t*-tests were utilized. If data was nonparametric, Wilcoxon Signed Rank Tests were utilized for statistical significance ($p < 0.050$). For practical significances, we utilized effect sizes (*d*-values 0 to 0.2, 0.2 to 0.5, and $d > 0.8$ for small, medium, and large practical significance, respectively).

**Entering Math Anxiety and Effect of Content Course**

To answer our first two research questions, we determined the range, mean($\mu$), standard deviation($\sigma_{\bar{x}}$), *p*-value (statistical significance), and *d*-value (practical significance) of math anxiety for the 160 preservice teachers before and after the content course. Results are given in Table 1.

**Table 1**
*Math Anxiety Levels Before and After Content Course*

| | | Entering preservice training | | | | After content course | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | *p*-value | *d*-value |
| Pre-service teachers | 160 | 68 | 128 | 82.91 | 13.94 | 49 | 112 | 78.48 | 12.39 | 0.000 | 0.34 |

Table 1 indicates that students entered their training with an average R-MANX score of 82.9 and exited the course with an average score of 78.5, a 4.4-point (5.3%) decrease. The standard deviation decreased from 13.9 to 12.4 after the course. The range in anxiety scores was large. For example, before the course, the lower quartile (low anxiety) spanned a narrow 5-point range, and the upper quartile extended over a much larger 36-point range (see Figure 1). After the course, the lower quartile increased to a 21-point range, and the upper quartile span decreased to a 26.5-point range. Additionally, the minimum score dropped 19 points, and the maximum score had decreased by 16 points. This freshmen-level content course had a statistically significant impact ($p < 0.05$) and a medium practical effect size (0.34).

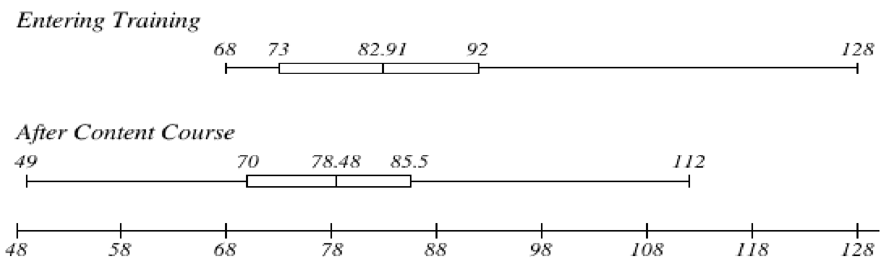


***Figure 1.*** Math anxiety data distribution.

**Teaching Style**

To address our third research question, the impact of teaching style, we compared Instructor A's (PBT) students to Instructor B's (Direct T) students (see Table 2). Only if there was a statistical significant difference ($p < 0.050$) did we investigate effect sizes (see Table 2). The mean score for the Direct T students increased from 75.1 to 76.6 (about 2%). In contrast, the scores for the PBT students lowered from 88.5 to 79.9 (nearly 10%). Attending a problem-based class led to a statistically and practically significant decrease in math anxiety, but attending a Direct T class did not.

**Table 2**
*Impact of Lecturer's Teaching Style on Math Anxiety Levels*

| | | Entering preservice training | | | | After content course | | | | Impact | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Points | *p-value* | *d-value* |
| PBT | 93 | 68 | 128 | 88.54 | 12.37 | 53 | 112 | 79.85 | 12.03 | -8.69 | 0.000 | 0.71 |
| Direct T | 67 | 48 | 107 | 75.09 | 12.61 | 49 | 106 | 76.58 | 12.80 | +1.49 | 0.335 | - |

A comparison of the mean anxiety score before the content course indicates that the students in the PBT classes started with higher anxiety than students in the Direct T classes. A Mann-Whitney Test comparing the Direct T and PBT students before the course indicated that there was a statistically significant difference ($p = 0.000$). This is not surprising because students were not pre-filtered into classes. The reductions shown in Table 2 for the PBT population may be due to the PBT population starting with substantially higher anxiety. That is, it may be easier to decrease anxiety in high anxiety students, no matter the teaching style. This question can be addressed by assessing the impact of the course (and teaching style) on (a) the high anxiety students in Direct T courses and (b) the low anxiety students in PBT courses. The median value of 82 for the total population was used to divide students into two groups: high and low anxiety populations. These analyses showed that (a) the high anxiety students in the Direct T courses ($n$ = 17) showed no statistically significant changes ($p$ = 0.394 via a Wilcoxon Signed Rank Test) and (b) the low anxiety students in the PBT courses ($n$ = 36) showed a 5-point decrease in the mean which was statistically significant ($p$ < 0.0001 via a paired $t$-test with a large effect size, $d$ = 1.03). These analyses indicate that the PBT teaching style does have a large impact for high and low anxiety students. In contrast, the direct teaching style had no significant impact on either high or low anxiety students.

**Disciplinary Majors**

Our expectation was that nonmath majors may start with a higher level of math anxiety, due to a possible lower level of content knowledge, and be impacted more positively by the content course as they gain knowledge to teach math. We defined math teaching (MATH-t) majors as both I-STEM ($n$ = 18) and MATH ($n$ = 2) majors because both will be certified to teach higher levels of math. Our expectation was that MATH-t majors would start with a lower level math anxiety and that the content course would reduce nonmath majors' math anxiety more than math majors. The analysis of math anxiety by math and nonmath majors is presented in Table 3.

**Table 3**
*Impact of Lecturer's Teaching Style on Math Anxiety Levels of Math and Nonmath Majors*

| | | Entering preservice training | | | | After content course | | | | Impact | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Points | $p$-value | $d$-value |
| Both instructors | | | | | | | | | | | | |
| MATH-t | 20 | 62 | 109 | 85.20 | 13.70 | 53 | 94 | 77.55 | 10.78 | -7.65 | 0.009 | 0.62 |
| OTHERS | 140 | 48 | 128 | 82.58 | 13.99 | 49 | 112 | 78.61 | 12.63 | -3.96 | 0.000 | 0.30 |
| PBT | | | | | | | | | | | | |
| MATH-t | 10 | 86 | 109 | 96.40 | 7.26 | 53 | 94 | 80.80 | 12.41 | -15.60 | 0.006 | 1.53 |
| OTHERS | 83 | 68 | 128 | 87.46 | 12.30 | 63 | 112 | 79.73 | 11.84 | -7.72 | 0.000 | 0.64 |
| Direct T | | | | | | | | | | | | |
| MATH-t | 10 | 62 | 90 | 74.00 | 8.04 | 63 | 88 | 74.30 | 8.23 | 0.30 | 0.838 | - |
| OTHERS | 57 | 48 | 107 | 75.28 | 13.15 | 49 | 106 | 76.98 | 13.64 | 1.70 | 0.274 | - |

The results for the total population (both instructors) show that both math and nonmath majors benefitted from attending the content course because both were statistically significant. The 20 math majors showed the greatest decrease in math anxiety, 7.7-points (nearly 9%). The practical significance was medium for the MATH-t majors and small for the nonmath students.

More significant differences by disciplinary major were apparent when accounting for teaching style. Instructor B (Direct T) had no statistically significant impact on either MATH-t or nonmath students. By contrast, Instructor A (PBT) had a large positive impact on MATH-t and nonmath majors, which was statistically significant with large to medium effect sizes.

**Education Majors**

The analysis of math anxiety by different education majors is presented in Table 4. A one-way ANOVA test, using the total population, resulted in a *p*-value of 0.344 indicating that the four groups were not statistically significantly different before the content course. Using the total population, all four education major groups showed decreases on anxiety of 4 to 6 points, three of which were statistically significant with medium practical significance. (The fourth group, DEAF, was close to significant with $p = 0.057$.)

More significant differences by education major were apparent when separating teaching style. All education majors in the PBT courses had statistically significant decreases in anxiety, which had medium to large effect sizes. The students in the Direct T courses had a substantially smaller impact, with predominately increases of anxiety. Only one subgroup (ELEM) had a statistically significant difference ($p = 0.049$), an increase in anxiety of 2.65 points.

The SPED group did appear to be unique in that anxiety reductions were observed for both Direct T and PBT classes (but with only the PBT group being statistically significant).

**Table 4**
*Impact of Lecturer's Teaching Style on Math Anxiety Levels of Different Education Majors*

| | | Entering preservice training | | | | After content course | | | | Impact | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Range | | $\mu$ | $\sigma_{\bar{x}}$ | Points | *p-value* | *d-value* |
| Both instructors | | | | | | | | | | | | |
| EACH | 36 | 63 | 123 | 85.22 | 14.56 | 58 | 112 | 81.03 | 13.68 | -4.19 | 0.008 | 0.30 |
| ELEM | 79 | 49 | 128 | 81.34 | 13.82 | 51 | 106 | 76.68 | 11.24 | -4.66 | 0.000 | 0.37 |
| DEAF | 23 | 57 | 109 | 86.00 | 13.87 | 63 | 107 | 82.04 | 13.96 | -3.83 | 0.057 | - |
| SPED | 22 | 48 | 103 | 81.50 | 13.25 | 49 | 94 | 75.77 | 11.39 | -5.73 | 0.004 | 0.46 |
| PBT | | | | | | | | | | | | |
| EACH | 24 | 69 | 123 | 90.25 | 13.90 | 63 | 112 | 82.04 | 13.55 | -8.21 | 0.000 | 0.60 |
| ELEM | 48 | 68 | 128 | 87.17 | 11.71 | 53 | 106 | 77.79 | 10.45 | -9.38 | 0.000 | 0.85 |
| DEAF | 10 | 74 | 109 | 91.50 | 11.77 | 70 | 107 | 84.30 | 16.05 | -5.00 | 0.009 | 0.51 |
| SPED | 11 | 71 | 103 | 88.09 | 10.84 | 66 | 93 | 78.64 | 10.39 | -9.45 | 0.004 | 0.89 |
| Direct T | | | | | | | | | | | | |
| EACH | 12 | 63 | 90 | 75.17 | 10.25 | 58 | 106 | 79.00 | 14.30 | 3.83 | 0.290 | - |
| ELEM | 31 | 49 | 100 | 72.32 | 11.96 | 51 | 101 | 74.97 | 12.34 | 2.65 | 0.049 | 0.22 |
| DEAF | 13 | 57 | 107 | 81.77 | 14.29 | 63 | 105 | 85.80 | 13.59 | 4.03 | 0.625 | - |
| SPED | 11 | 48 | 89 | 74.91 | 12.49 | 49 | 94 | 72.91 | 12.11 | -2.0 | 0.286 | - |

## Summary

Students entering their training had an R-MANX math anxiety level of 82.9. A required math content course (for educators) was useful in reducing math anxiety. Reductions in math anxiety were observed across education and disciplinary majors. Teaching style had a large beneficial impact on math anxiety, with a PBL style exhibiting statistically significant decreases and medium to large practical differences. In contrast, a direct teaching style had either no impact or a detrimental impact on anxiety. All subgroups were beneficially impacted by a PBL teaching style, but only SPED majors were beneficially impacted by a direct teaching style (not statistically significant).

## Discussion

We agree with Mercer et al. (1996) that a single instructional method is seldom effective for all students; however, in this study, a PBL-centric teaching style profoundly decreased math anxiety in an education contextualized math content course. This has substantial implications for both the implementation and impact of I-STEM methods as well as how we train T&E teachers. In general, our T&E teacher preparation programs have limited math, science, and engineering (each having important math contexts). This lower emphasis on math likely results in higher math anxious T&E teachers and limited implementation of PBL methods, and it certainly limits how much engineering (vs. technology) can be effectively addressed in classes. Litowitz (2014) found that 75% of our T&E teacher preparation programs required only lower level math courses. Additionally, Litowitz (2014) found only one program with a required contextualized (engineering) math course. A lower emphasis on math has also been evident in our certified teachers. When investigating familiarity with the grade level of mathematics standards, Flowers and Rose (2014) found that T&E teachers were (a) only accurate 40% of the time and (b) off by two or more grade levels 30% of the time. Additionally, mathematics is also not represented substantially in field research. Of the 97 papers published in this journal from spring 2007 through spring 2016, only seven had the word *mathematics* in the title. Strimel and Grubbs (2016) also discussed several of these observations, as well as other observations, when suggesting a larger emphasis on engineering in the field. Because I-STEM teaching utilizes design-centric PBL methods, this study indicates that I-STEM activities may lower math anxiety and therefore increase mathematical skills in both K–12 students and preservice T&E teachers. If T&E teacher preparation programs generally required more education-centric or contextualized math courses (especially utilizing a PBL teaching style), then I-STEM (or Engineering) methods might be more effectively be implemented. Burghardt, Hecht, Russo, Lauckhardt, and Hacker (2010) also suggests that mathematically integrated PBL-centric I-STEM methods be modeled in preservice T&E teacher programs.[1]

## Future Work

Questions that remain include inquiries on the longevity of this impact as well as extending the study to specific preservice or in-service elementary or secondary teachers in T&E or STEM and extending the study to other institutions.

[1] It should be noted that technology education has worked on integrated curricula math projects (LaPorte & Sanders, 1993; Satchwell & Loepp, 2002).

## References

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Sciences*, *11*(5), 181–185. doi:10.1111/1467-8721.00196

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Barrows, H. (2002). Is it truly possible to have such a thing as dPBL? *Distance Education*, *23*(1), 119–122. doi:10.1080/01587910220124026

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5) 1860–1863. doi:10.1073/pnas.0910967107

Brady, P., & Bowd, A. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among pre-service education students. *Teachers and Teaching: Theory and Practice*, *11*(1), 37–46. doi:10.1080/1354060042000337084

Brown, A. B., Westenskow, A., & Moyer-Packenham, P. S. (2011). Elementary pre-service teachers: Can they experience mathematics teaching anxiety without having mathematics anxiety? *Issues in the Undergraduate Mathematics Preparation of School Teachers: The Journal*, 5. Retrieved from http://www.k-12prep.math.ttu.edu/journal/5.attributes/brown01/article.pdf

Buhlman, B. J., & Young, D. M. (1982). On the transmission of mathematics anxiety. *Arithmetic Teacher*, *30*(3), 55–56.

Burghardt, M. D., Hecht, D., Russo, M., Lauckhardt, J., & Hacker, M. (2010). A study of mathematics infusion in middle school technology education classes. *Journal of Technology Education*, *22*(1), 58–74. doi:10.21061/jte.v22i1.a.4

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science. *School Science and Mathematics*, *106*(4), 173–180. doi:10.1111/j.1949-8594.2006.tb18073.x

Burton, L. (1998). Undergraduate engineering education: Teaching, learning, assessing—A symbiosis. *Engineering Science and Education Journal*, *7*(4), 158–160. doi:10.1049/esej:19980404

Cornell, C. (1999). "I hate math! I couldn't learn it, and I can't teach it!" *Childhood Education*, *75*(4), 225–231. doi:10.1080/00094056.1999.10522022

Finlayson, M. (2014). Addressing math anxiety in the classroom. *Improving Schools*, *17*(1), 99–115. doi:10.1177/1365480214521457

Flowers, J., & Rose, M. A. (2014). Mathematics in technology & engineering education: Judgments of grade-level appropriateness. *Journal of Technology Education*, *25*(2), 18–34. doi:10.21061/jte.v25i2.a.2

Frank, M. L. (1990). What myths about mathematics are held and conveyed by teachers? *Arithmetic Teacher*, *37*(5), 10–12.

Geist, E. (2010). The anti-anxiety curriculum: Combating math anxiety in the classroom. *Journal of Instructional Psychology*, *37*(1), 24–31.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569–582. doi:10.1037/0022-0663.76.4.569

Gresham, G. (2007). A study of mathematics anxiety in pre-service teachers. *Early Childhood Education Journal*, *35*(2), 181–188. doi:10.1007/s10643-007-0174-7

Gresham, G. (2008). Mathematics anxiety and mathematics teacher efficacy in elementary pre-service teachers. *Teaching Education*, *19*(3), 171–184. doi:10.1080/10476210802250133

Harper, N. W., & Daane, C. J. (1998). Causes and reductions of math anxiety in preservice elementary teachers. *Action in Teacher Education*, *19*(4), 29–38. doi:10.1080/01626620.1998.10462889

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46. doi:10.2307/749455

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *The Interdisciplinary Journal of Problem-Based Learning*, *1*(1), 21–39. doi:10.7771/1541-5015.1004

Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2009). STEM integration in K–12 education: Status, prospects, and an agenda for research. Washington DC: National Academies Press. doi:10.17226/18612

Isiksal, M., Curran, J. M., Koc, Y., & Askun, C. S. (2009). Mathematics anxiety and mathematical self-concept: Considerations in preparing elementary-school teachers. *Social Behavior and Personality*, *37*(5), 631–644. doi:10.2224/sbp.2009.37.5.631

Jackson, C. D., & Leffingwell, R. J. (1999). The role of instructors in creating math anxiety in students from Kindergarten through college. *Mathematics Teacher*, *92*(7), 583–586.

Kelly, W. P., & Tomhave, W. K. (1985). A study of math anxiety/math avoidance in preservice elementary teachers. *Arithmetic Teacher*, *32*(5), 51–53.

LaPorte, J., & Sanders, M. (1993). The T/S/M Integration Project: Integrating technology, science, and mathematics in the middle school. *The Technology Teacher*, *52*(6), 17–21.

Liao, C. (2016). From interdisciplinary to transdisciplinary: An arts-integrated approach to STEAM education. *Art Education*, *69*(6), 44–49. doi:10.1080/00043125.2016.1224873

Litowitz, L. S. (2014). A curricular analysis of undergraduate technology & engineering teacher preparation programs in the United States. *Journal of*

*Technology Education*, *25*(2), 73–84. doi:10.21061/jte.v25i2.a.5

Mercer, C. D., Lane, H. B., Jordan, L., Allsopp, D. H., & Eisele, M. R. (1996). Empowering teachers and students with instructional choices in inclusive settings. *Remedial and Special Education*, *17*(4), 226–236. doi:10.1177/074193259601700405

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. doi:10.1037/h0033456

Sanders, M. (2009). STEM, STEM education, STEMmania. *The Technology Teacher*, *68*(4), 20–26.

Satchwell, R. E., & Loepp. F. L. (2002). Designing and implementing an integrated mathematics, science, and technology curriculum for the middle school. *Journal of Industrial Teacher Education*, *39*(3), 41–66.

Strimel. G., & Grubbs, M. E. (2016). Positioning technology and engineering education as a key force in STEM education. *Journal of Technology Education*, 27(2), 21–36. doi:10.21061/jte.v27i2.a.2

Strobel, J., & van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-Based Learning*, *3*(1), 44–58. doi:10.7771/1541-5015.1046

Swars, S. L., Daane, C. J., & Giesen, J. (2006). Mathematics anxiety and mathematics teacher efficacy: What is the relationship in elementary preservice teachers? *School Science and Mathematics*, *106*(7), 306–315. doi:10.1111/j.1949-8594.2006.tb17921.x

Swetman, B., Munday, R., & Windham, R. (1993). Math-anxious teachers: Breaking the cycle. *College Student Journal*, 22(4), 421–427.

Teague, P. T., & Austin-Martin, G. (1981, February). *Effects of a mathematics methods course on prospective elementary school teachers' math attitudes*, *math anxiety*, *and teaching performance.* Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

Trujillo, K. M., & Hadfield, O. D. (1999). Tracing the roots of mathematics anxiety through in-depth interviews with preservice elementary teachers. *College Student Journal*, *33*(2), 219–232.

Uusimaki, L., & Nason, R. (2004). Causes underlying pre-service teachers' negative beliefs and anxieties about mathematics. In *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 369–376). Retrieved from https://www.emis.de/proceedings/PME28/RR/RR141_Uusimaki.pdf

Vinson, B. M. (2001). A comparison of preservice teachers' mathematics anxiety before and after a methods class emphasizing manipulatives. *Early Childhood Education Journal*, *29*(2), 89–94. doi:10.1023/A:1012568711257

Virginia Tech. (2017). Integrative STEM education. Retrieved from http://www.soe.vt.edu/istemed/

Wadlington, E., & Wadlington, P. L. (2008). Helping students with mathematical disabilities to succeed. *Preventing School Failure*, *53*(1), 2–7. doi:10.3200/PSFL.53.1.2-7

Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, *3*(1), 12–43. doi:10.7771/1541-5015.1061

Wells, J. G. (2016). PIRPOSAL model of integrative STEM education: Conceptual and pedagogical framework for classroom implementation. *Technology and Engineering Teacher*, *75*(6), 12–19.

Zientek, L. R., Thompson, B., & Yetkiner, Z. E. (2010). Characterizing the mathematics anxiety literature using confidence intervals as a literature review mechanism. *Journal of Educational Research*, *103*(6), 424–438. doi:10.1080/00220670903383093

## About the Authors

**Suriza Van der Sandt** (vandersa@tcnj.edu) is Associate Professor in the Department of Mathematics and Statistics at The College of New Jersey.

**Steve O'Brien** (obriens@tcnj.edu) is Associate Professor and Chair of the Department of Technological Studies at The College of New Jersey.

# The Need, Development, and Validation of the Innovation Test Instrument

*Jacob Wheadon, Geoff A. Wright, Richard E. West, & Paul Skaggs*

**Abstract**

This study discusses the need, development, and validation of the Innovation Test Instrument (ITI). This article outlines how the researchers identified the content domain of the assessment and created test items. Then, it describes initial validation testing of the instrument. The findings suggest that the ITI is a good first step in creating an innovation assessment because it is more inclusive of both divergent and convergent thinking. In comparison, past innovation assessments have only assessed either divergence or convergence. The ITI still needs further validation and improvement to make strong claims about its ability to determine the effectiveness of an innovation course.

*Keywords:* Innovation, assessment, validity, creativity

This article is based on the Master's Degree Thesis Wheadon, J. D. (2012). *Development and initial validation of an innovation assessment* (Master's thesis, Brigham Young University). Retrieved from http://scholarsarchive.byu.edu/etd/3326/

## The Need for Innovation

In industry and education, there is an increasing push for organizations and individuals to be more innovative (Fagerberg, 1999; Wagner, 2010). Rapid technological change has created the need for organizations and individuals to adapt quickly (Christensen & Eyring, 2011). Christensen (1997) describes how disruptive innovations fundamentally change markets and require new ways of thinking for organizations to adapt and survive. He describes how individuals in organizations need to think differently in order to compete in today's marketplace. Because of the rapid rate of technological change that is occurring today, disruptive innovations are changing markets even faster than in the past. This has led to a greater need for people to cultivate innovation skills.

Innovation skills are also needed to create job growth. Various economies have made claims and refocused their industries to further promote and harness innovation. The European Union (EU) reported that "the central aim of the EU 2020 strategy is to put Europe's economies onto a high and sustainable growth path. To this end, Europe will have to strengthen its innovative potential and use its resources in the best possible way" (European Commission, 2011, p. 2). Similarly, the Federal Bureau of Business and Economics of India stated: "In the ever-changing world, innovation is the only key which can sustain long-run growth of the country . . . innovation [provides] competitive advantage" (National Portal of India, 2014). In the United States, innovation had been reported as the de facto source of job creation since the 20th century (Drucker, 1985). Drucker (1985), Wagner (2012), Former President Barack Obama (The White House: President Barack Obama, 2011), and Friedman and Mandelbaum (2011), among others, have all advocated for the growth and development and the need for people and organizations to be more innovation—to be globally competitive and marketable.

### The Need to Teach Innovation

Many of these calls for increased innovation have mentioned the need for schools to teach students to be more innovative (Friedman & Mandelbaum, 2011; Wagner, 2010; Wagner 2012). They have said that for American students to remain competitive in a global market and be able to adapt to a constantly shifting playing field, they need to become innovators. Schools need to teach students the skills and behaviors of great innovators (Wagner, 2010).

In a recent study, Dyer, Gregersen, and Christensen (2011) identified the common behaviors that many of today's leading innovators share. By studying innovators' behaviors, they found that people who want to be better innovators can learn and practice behaviors that will help them create innovations. Dyer et al. give educators a set of teachable skills that students can learn to perform. They claimed that although some people might have a natural propensity for innovation, anyone can learn to be more innovative.

With the knowledge that innovation can be taught, some schools, consulting firms, and corporations have begun teaching innovation. Well-known examples include the Hasso Plattner Institute of Design at Stanford University (d.school; 2017; Stanford Graduate School of Business, 2017), IDEO (IDEO, 2017; Kelly, 2005), and Innosight (Innosight, 2011), who have all reported the great value and impact of their teaching about innovation.

The College of Engineering and Technology at Brigham Young University (BYU) has a three-fold mission statement, and innovation is central to that mission. Consequently, a faculty committee was created with the goal of developing a course to teach innovation. The course curriculum uses an active learning pedagogy, teaches students about the need for innovation, and engages them in various activities during which they practice and develop divergent and convergent thinking skills and behaviors (Howell, Skaggs, & Fry, 2010). The course is currently known as the Innovation Bootcamp, and its curriculum is focused on teaching an innovation model that promotes idea finding, idea shaping, idea defining, idea refining, and idea communicating.

## The Need to Assess Innovation Teaching

The Innovation Bootcamp in various forms has been taught in the College of Engineering and Technology since 2008. The course consistently receives very positive student feedback on end of term evaluations. In addition, informal assessments asking students to report on their level of interest and ability in using innovation pre- and post-course suggested that the course was having a positive impact. However, because the informal assessments were not initially designed with the intent of a longitudinal study of testing student innovative ability, the researchers believed that an assessment should be developed to ensure that course learning outcomes were being met. In addition, they believed that an innovation assessment such as this would prove to be of significance to others interested in assessing innovative ability.

## Current Innovation Assessments

Tyler Lewis's (2011) thesis, *Creativity and Innovation: A Comparative Analysis of Assessment Measures for the Domains of Technology, Engineering, and Business*, analyzed various innovation and creativity assessments and measures. His findings suggested that innovation was either being measured in terms of creativity or divergent thinking (i.e., creativity tests often focused directly on divergent thinking; Houtz & Krug, 1995). Other creativity tests measure different aspects of divergent thinking, such as flexibility (Torrance, 1963), fluency (Houtz & Krug, 1995; Torrance, 1963), and originality (Houtz & Krug, 1995; Torrance, 1963), or focus on the environment for promoting innovation or focus on the end or implementation of the product (convergent thinking). For example, measures in Radosevic and Mickiewicz (2003) evaluated the success of innovation programs in terms of financial outputs, such as sales of a product or an increase in profits during or after the introduction of an innovation course or program. However, the measures that Lewis (2011) suggested would not be accurate for measuring people's innovative abilities.

The instructors of the Innovation Bootcamp implemented various measures such as the Torrance Test of Creative Thinking (TTCT) but found that these types of assessments, as Lewis (2011) had postulated, only measured the divergent thinking (creativity) part of innovation. Still needing a innovation assessment that would assess a person's innovative ability, the researchers decided to develop their own assessment to measure both divergent and convergent thinking.

## Methodology

The faculty members involved with the development of the Innovation Bootcamp visited various recognized innovation institutions such as Innosight, IDEO, and Stanford's d.school, among others, and completed a very comprehensive literature review of innovation principles, methods, and processes. They ultimately identified five common themes in the innovation research, which they used as the primary content stands for the Innovation Bootcamp. The five content strands, or "phases of innovation," are: idea finding, idea sharping, idea defining, idea refining, and idea communicating (see Figure 1).

The focus of Idea Finding is on helping students to be able to identify opportunities for innovation (some call this the problem-finding phase). The research on innovation suggests a wide variety of tools to help people identify or find innovation opportunities. The Bootcamp focused on teaching students three such tools in the areas of observing, experiencing, and inquiring.

The purpose of the second phase, Idea Shaping, is to help students organize, simplify, and clarify the results from their observations, experiences, or inquiries from the Idea Finding phase.

The third phase, Idea Defining, helps the students start to solve the problem that they identified from the previous two phases. Some researchers define this phase as brainstorming; however, it is more than simply generating a variety of options. This phase is concerned with associating and connecting ideas that may seem unrelated with the intent of forming ideas that are highly useful and novel.

The fourth phase is Idea Refining. During this phase, students are taught how to visualize, validate, and iterate the potential solutions that they generated in the previous phases. Other innovation researchers might connect or associate this phase with prototyping. However, the researchers at the Innovation Bootcamp believe that this phase is more than prototyping because it also promotes the need to decide the validity and value of the solution. This phases also stresses the idea of rapid prototyping in any format, from basic card stock and sketches to wire mockups and photo manipulations. The Idea Refining phase uses the motto of "anything that can quickly communicate your idea" to prompt students.

The final phase, Idea Communicating, teaches students how to communicate their solutions and ideas to others. This phase is taught by providing examples and rationale showing that presentations are insufficient to communicate an idea; there is a need to show, demonstrate, and describe within a context or situation. Meaning that a solution must be presented within the context of how the solution will fulfill the demand or problem.

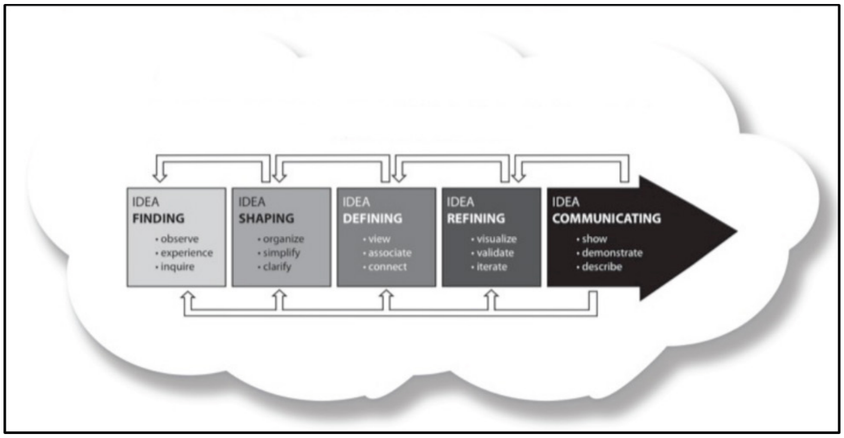


***Figure 1.*** BYU Innovation Bootcamp model.

The five phases were used to organize the learning outcomes for the course, which guided the creation of the assessment. The learning outcomes were organized into four parts: opportunity recognition (Phases 1 and 2 of the innovation curriculum), ideation (Phase 3), idea refining (Phase 4), and communication (Phase 5). The four learning outcomes were used to create a two-way chart that was used to organize what needed to be measured in the assessment. The two-way chart, called a table of specifications (Miller, Linn, & Gronlund, 2009), is a common tool used in the development of tests, assessments, and curriculum development (Table 1) in which content strands are listed on one axis and cognitive processes are listed on the other axis. Bloom's Revised Taxonomy was the foundation for the cognitive processes in the Innovation Test Instrument (ITI; Anderson & Krathwohl, 2001). Bloom's Revised Taxonomy was used because it is a well-known and respected list of cognitive processes, and this list aligned with the course's learning outcomes. The course's learning outcomes focus on application by inviting students to apply what they are learning, so two test items were created to meet this demand. Because the course teaches students how to analyze opportunities for innovation in the various problem-spotting activities, two test questions were created to align with this cognitive process. The cognitive process of evaluation was also a key element of the course's learning outcomes; therefore, two test questions were related to this process. In these two questions, students were required to justify their decisions for the newly designed innovation. Finally, in the cognitive process of creation, the desired outcome was to assess an individual's ability to prototype an idea. A prototype is defined as a strong visual manifestation. Consequently, in the two test questions related to creation, students were required to draw and annotate the new product, system, or service that they came up with.

**Table 1**
*Table of Specifications*

| | Remember | Understand | Apply | Analyze | Evaluate | Create |
|---|---|---|---|---|---|---|
| Opportunity recognition | | | | 2 | | |
| Ideation | | | | | | 2 |
| Idea refining | | | | | 2 | |
| Communication | | | 2 | | | |

The table of specifications (see Table 1) shows the number of items created for each learning outcome. Ultimately, there were assessment items made in the apply, analyze, evaluate, and create cognitive-process areas.

The first item type corresponded with the first learning outcome and tested students' ability to find problems using a photo-identifying activity. In this activity students were asked to identify as many areas or behaviors that were problematic. Students were graded on how many problems they were able to identify within a specified amount of time. Higher scores were awarded to those who identified more novel problems (novelty was measured using student response frequency).

In the second item type, students were given a problem statement (i.e., bike seats get wet) and were asked to write out as many solutions as they could within a specified amount of time. Higher points were again awarded for more novel but feasible answers. The TTCT uses a similar grading scheme (Torrance, 1963).

The third item type assessed the students' ability to evaluate ideas by presenting a series of possible solutions to a given problem and asking them to rank order the solutions from best to worst. Their rankings should have been based on the definition of innovation used by the Innovation Bootcamp: original and useful ideas that can be implemented successfully. The student responses were compared with the responses of four technology and engineering professors who have significant experience in innovation research and industry. To ensure interrater reliability, the responses of the professors were compared and analyzed prior to comparing them with the student responses.

The final item type assessed the students' abilities to effectively communicate their ideas to others. This item required students to write out a pitch for the innovative solution that they ranked the highest on the previous ranking question. The pitch was limited to 700 characters, which meant that it had to be concise. The grading of the pitch was based on conciseness and effective communication of the value of the solution.

The final item was graded by two raters using the provided rubric. Raters were trained on how to use the rubric and then graded five questions. They graded preselected responses that were considered by the researchers to be good, mid-grade, and poor in order to ensure that the raters could be reliable at different levels of performance. The raters discussed any areas in which they disagreed. After grading the first five responses and their subsequent discussion, the raters graded five more responses and then discussed the scores. This process continued until raters achieved agreement, which was defined as a correlation greater than 0.75 because an interrater reliability above 0.75 is considered "excellent" (Cicchetti, 1994, p. 286). After the raters graded all responses, interrater reliability was estimated for all scores.

**Testing Procedures**

An initial pilot version of the test was first administered during the fall semester (2012) of the Innovation Bootcamp course. It was administered to three sections of the course, which had 20 students in each section ($n$ = 60). The pilot version was done to help with initial test form equivalence and instrument validity. Following the initial pilot implementation, the results were analyzed, and the test was revised. The revised version of the test was then administered during the winter semester of the course to five sections of the Innovation Bootcamp ($n$ = 100). Students were told that the test was a contest and that the top scores would receive a cash prize. The extrinsic motivation of a cash prize was added based on the results from the pilot test, which suggested that we needed to ensure students were motivated to do their best on their test to ensure maximal performance.

**Revisions to the ITI after the initial test.** After the initial test, the results were analyzed and revisions to the ITI were made in order to improve the test. The biggest problem with the initial test was that the subjects did not achieve maximal performance. Few of the subjects finished the test, and others quickly went through the items without giving much thought to them. This likely happened for a couple of reasons. The first reason is test fatigue. Subjects' performance dropped off significantly the longer they spent on the test. This was remedied by making the test shorter. The original length of the test was longer so that there would be a larger item bank for future testing. This proved infeasible for this study because the subjects could not maintain concentration over the large number of items.

The second reason for inadequate performance was that the stakes were not sufficiently high to prompt maximal performance. In order to resolve this issue, the second round of testing was done as a competition. Cash prizes were offered to subjects with the highest test scores.

Fixing these two problems with the test strengthened evidence of construct validity. Problems with fatigue and lack of incentive hurt the construct validity of the test. Problems in the test procedure affected scores enough that they did not accurately describe a person's ability to perform the tasks. By fixing these problems, a stronger claim of construct-related evidence can be made.

**Test form equivalence.** Because a major part of this study was to create equivalent forms that can be used for pre- and post-testing, two forms of the test were created and given to the students at the same time. To find the forms equivalent, corresponding items should have similar means and standard deviations for the same group of test subjects. Also, student rankings by total score should be the same for both forms of the test.

## Results

### Overall Results for the Initial Test

The initial (or pilot) test was given to the three sections of the Innovation Bootcamp in the fall semester. The participants were split into two groups. Half of the students from each class were put into Group A, and half were put into Group B. Table 2 lists the participant scores and the means and standard deviations for the groups.

**Table 2**
*Summary of Overall Scores for the Initial Test*

| | Group A | | | Group B | | |
|---|---|---|---|---|---|---|
| | Overall | Form 1[a] | Form 2 | Overall | Form 1 | Form 2[a] |
| Mean | 75.83 | 44.92 | 30.92 | 98.17 | 46.33 | 51.83 |
| *SD* | 36.95 | 15.67 | 21.88 | 43.58 | 21.60 | 23.60 |
| Correlation | .93 | | | .86 | | |

[a] Indicates which form was taken first by each group (Group A started with Form 1, and Group B started with Form 2).

These data show that scores declined as test time increased, meaning that, regardless of the test form, averaged scores were lower on the second test form. For example, Group A's mean scores decreased from 44.92 to 30.92, which was similar to Group B's decrease from 51.83 to 46.33. Although the decline was lower in Group B, because both groups experienced a decline, this was attributed to (a) test fatigue and (b) lack of incentive.

Observation showed that the subjects became fatigued because of the length of the test and the number of items. For example, many of the subjects did not attempt to complete later items on the second form. Because of this finding, the test was modified into a significantly shorter version. Originally, each form of the test was going to have two items of each type; however, only one item of each type was included on each form of the revised version to reduce test fatigue.

Another limitation of the results is that many of the students failed to achieve maximal performance on the test items because they were not interested enough in completing the test (not enough incentive). Some subjects skipped essay questions or answered them with only a few words, which was problematic because the test was designed to score participants based on subjects' maximal performance of cognitive tasks. In the initial trial of the test, stakes were not high enough to prompt maximal performance. Consequently, incentives were offered for high performance on the revised version of the test.

**Analysis of Individual Items**

Analysis of the scores and responses for individual items were used to gather evidence of validity and to find ways to improve the items for future tests. Even though the initial test's issues of length and test fatigue limited what could be learned from these results, there were still important things shown. Some of the items did not perform as expected and were revised for the second round of testing. The problem-finding items did not generate a large enough variety of responses and were modified. Also, the communication items needed better instructions and were modified to help the subjects understand better what was expected of them.

**Analysis of Problem-Finding Items**

In the problem-finding items, subjects tried to identify problems from photographs provided in the test. A rater counted all of the responses to find out which responses were more common than others. Figures 1–4 show the pictures used in each item.

***Figure 1.*** Photograph from the man on couch problem-finding item.

***Figure 2.*** Photograph from the leaky drain problem-finding item.

***Figure 3.*** Photograph from the printer problem-finding item.

***Figure 4.*** Photograph from the street cracks problem-finding item.

The mean scores and standard deviations are shown in Table 3, which includes the overall means and standard deviations as well as the means and standard deviations for the two test groups.

**Table 3**
*Summary of Statistics for Problem-Finding Items*

| | Overall | | Group A | | Group B | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Man on couch | 7.75 | 3.94 | 9.17 | 4.47 | 6.33 | 2.66 |
| Leaky drain | 7.88 | 5.24 | 8.17 | 6.15 | 7.58 | 4.11 |
| Printer | 7.33 | 5.91 | 6.58 | 5.68 | 11.08 | 6.78 |
| Street cracks | 6.71 | 5.59 | 5.75 | 5.83 | 7.33 | 5.47 |

These statistics show that there was a significant order effect. The subjects tended to perform better on items that they completed earlier in the test. This makes establishing equivalence between the items difficult because it is unknown whether the change in scores was a result of those items being more difficult or a result of the order in which the subjects completed the items. Notwithstanding the order effect, some claims can be made about the difficulty of the items. Both groups scored higher on the printer item than the street cracks item. Because these items were placed in the same section of the test, this difference can likely be attributed to difficulty of the items. The other scores were inconclusive. Even though the man on couch and leaky drain items were in the same section of the test, Group A performed better on the man on couch item, and Group B performed better on the leaky drain item. The man on couch and street cracks items showed less divergence in their responses. This led to the decision to test different photographs in the second round of testing. In this initial test, problem-finding photographs were taken of specific problems similar to the ones that students identify in the Innovation Bootcamp; however, in the revised version, the problem-finding items had pictures that were taken of scenes from a home without focusing on specific problems. It was hoped that these photographs would give subjects the opportunity to identify a wider range of problems and that having to identify problems from a broader scene would be closer to the experience of problem finding that students face in the Innovation Bootcamp and that innovators face in real-world practice.

**Analysis of Solution Items**

The solution items gave subjects problem statements and asked them to generate as many solutions as they could. The scoring of these items followed a similar procedure to the problem-finding items. Students received points for the solutions that they generate, and more points were awarded for novel (less common) responses.

The responses show that some of the items gave the subjects greater opportunities for different answers than others. The bakery item (i.e., a local supermarket has to discount their leftover baked goods after they are a day old) performed particularly poorly in this regard. It did not generate a very large number of different responses from the subjects. The garbage liner (i.e., garbage can liners often slip down inside of the cans when they are full of garbage) item performed best, followed by the headphone item (i.e., headphone wires get tangled in people's pockets), and then the corner-cutting item (i.e., people often cut across the lawn in places around campus, which leaves ugly dead patches in the grass). Other than the bakery item, these items garnered more responses than the problem-finding items. Table 4 shows the overall means and standard deviations as well as the means and standard deviations for the two test groups.

**Table 4**
*Summary of Statistics for Solution Items*

| | Overall | | Group A | | Group B | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Garbage liner | 7.33 | 5.91 | 5.50 | 2.25 | 9.17 | 7.61 |
| Headphone | 6.71 | 5.59 | 5.83 | 3.08 | 7.58 | 7.17 |
| Bakery | 5.71 | 4.25 | 4.50 | 3.75 | 6.92 | 4.37 |
| Corner cutting | 9.88 | 8.91 | 5.33 | 4.17 | 14.42 | 10.00 |

As with the problem-finding items, it is difficult to determine item equivalence based on the data shown here because of the order effect, which is attributed to test fatigue. These data show that for both groups, the bakery item was the most difficult. The other scores do not conclusively describe the equivalence of the other items.

The data from the solution items show that they performed better than the problem-finding items. In most of the items, the subjects gave a larger number of different responses than in the problem-finding items. Thus, the garbage liner and headphone items were chosen for more testing (to be used in the second round) because their means were closer than the others and because they had a large number of different responses.

**Analysis of Ranking Items**

The ranking items gave subjects a problem statement and four potential solutions. Participants ranked solutions using the Innovation Bootcamp's definition of innovation: original and useful ideas implemented successfully. Prior to administering the test, the ranking items were given to four engineering and technology professors. Their rankings were used to create a key to grade the students' scores by summing the point values from their rankings and then ranking the totals. Table 5 shows the overall and group means and standard deviations for the ranking items.

**Table 5**
*Summary of Statistics for Ranking Items*

| | Overall | | Group A | | Group B | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Bike seats | 4.92 | 3.08 | 5.58 | 2.98 | 4.25 | 3.03 |
| Toilets | 6.71 | 2.78 | 6.42 | 2.87 | 7.00 | 2.65 |
| Lawnmowers | 3.92 | 2.83 | 3.67 | 3.27 | 4.17 | 2.27 |
| Outlets | 2.88 | 2.11 | 3.00 | 2.24 | 2.75 | 1.96 |

The data show that the outlet item is more difficult than the other items because both groups did significantly worse on it than on the other three items. The lawnmower item also appears to have scored much lower, but in Group B, the lawnmower item scored close to the bike seat. Group A and the overall scores for the lawnmower item were lower. Because of this, the bike seat and toilet items were chosen to be retested in the revised test.

**Analysis of Communication Items**

The communication items followed the ranking items in the assessment. The communication items asked the subjects to create a pitch for the innovation that they ranked highest on the second ranking item. They were asked to create a convincing pitch that would persuade others to adopt the innovation that they chose. Table 6 shows the overall and group statistics for the communication items from each form of the instrument.

**Table 6**
*Summary of Statistics for Communication Items*

| | Overall | | Group A | | Group B | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Form 1 item | 4.33 | 3.57 | 4.25 | 3.42 | 4.42 | 3.71 |
| Form 2 item | 3.63 | 3.84 | 2.08 | 3.28 | 5.17 | 3.74 |

These data show that subjects in both groups performed poorly on both of the items. Although a total of 12 points were possible on the items, the means of the responses were less than half of that. A few problems with the items were observed when looking at individual responses.

The first problem was that many of the subjects gave very limited responses to these items. It appeared that the subjects did not care enough about the test to go through the effort of constructing a good response to this item. Also, many subjections did not finish the item. The researchers attempted to remedy this problem in the second round of testing by making the second round a competition with prizes for those with the highest scores on the test.

The second problem was that most subjects wrote the pitch as if the raters already understood the problem statement and the solutions. It was difficult for them to write about the problem and how the innovation fixed it when they were given both the problem and the solution. For this reason, in the revised version of the test, communicate questions were tied to the solution questions rather than the ranking questions. After the students generated their solutions from the given problem statement, the communication item was placed next so that students could explain the benefits of the innovation that they came up with rather than the innovation that they were given.

The third problem was that subjects did not always understand what they were supposed to write in the pitch. Some subjects described their rationale for choosing one of the responses over the others. Others failed to mention what the problem was or how their choice would solve that problem. To remedy this issue, clearer instructions were created for this item.

One aspect of these items that worked well was their rating. Using the grading rubrics, the raters scored the items with high reliability levels: 0.94 for the item from Form 1 and 0.97 from Form 2. Cicchetti (1994) said that reliability scores above 0.80 are considered "nearly perfect." This high reliability could be due to the training procedure explained in the methods section above but is also likely a result of so many of the responses being poor (raters easily agreed on responses that were severely lacking).

**Overall Results for the Revised Test**

The revised test was administered to 100 students in five sections of the Innovation Bootcamp. They were incentivized with cash prizes for the top 15 scores. To reduce test fatigue, the revised test also had half the number of questions that the initial pilot version did. The results show that having a shorter test with an incentive increased performance (see Table 7) and consistency—making the comparisons between items more helpful.

**Table 7**
*Summary of Scores for the Revised Test*

| | Group C | | | Group D | | |
|---|---|---|---|---|---|---|
| | Overall | Form 1[a] | Form 2 | Overall | Form 1 | Form 2[a] |
| Mean | 69.45 | 35.15 | 34.30 | 73.74 | 35.26 | 38.47 |
| *SD* | 17.95 | 9.74 | 9.81 | 21.28 | 9.76 | 13.28 |
| Correlation | .69 | | | .70 | | |

[a] Indicates which form was taken first by each group (Group C started with Form 1, and Group D started with Form 2).

**Results for Problem-Finding Items**

The problem-finding items on the revised version of the test used the same format as the initial version but with different pictures with a broader focus that the original pictures. The pictures used in the revised version of the test are shown in Figures 5 and 6.

***Figure 5.*** Photograph from the garage problem-finding item.

***Figure 6.*** Photograph for the bedroom problem-finding item.

The response counts revealed that the new problem-finding items garnered a much larger variation in the responses. The subjects gave many more and varied responses to the items than they did for the initial test. The mean scores and standard deviations of the problem-finding items are shown in Table 8. The table shows the overall means and standard deviations as well as the means and standard deviations for the two test groups.

**Table 8**
*Summary of Statistics for Problem-Finding Items*

| | Overall | | Group C | | Group D | | Item correlation |
|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | |
| Garage | 13.00 | 6.14 | 12.95 | 4.98 | 13.05 | 7.15 | 0.68 |
| Bedroom | 9.69 | 5.89 | 9.20 | 4.12 | 10.21 | 7.27 | |

These data show that the revised version of the test had a smaller order effect than the initial version. With the reduced order effect, the equivalence of the items could be studied. The difference between the means of the two items suggests that they cannot be considered equivalent. There appeared to be more problems to find in the garage item than in the bedroom item. In order to create two items that are more equivalent, more pictures should be tested and analyzed.

**Results for Solution Items**

The solution items on the revised test remained unchanged from the original test items. They appeared to be working well in the first test, but it was unclear how equivalent they were because of the order effect, so they were tested again in the revised test. The mean scores and standard deviations for the solution items are shown in Table 9.

**Table 9**
*Summary of Statistics for Solution Items*

| | Overall | | Group C | | Group D | | Item correlation |
|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | |
| Headphones | 8.95 | 4.85 | 8.95 | 5.04 | 8.95 | 4.64 | 0.46 |
| Garbage liner | 11.15 | 6.24 | 9.60 | 5.67 | 12.79 | 6.39 | |

The data in this table show that the order effect was also reduced for the solution items. The second round of testing gave a clearer view of the equivalence of the items. Because of the large difference in the means, the headphone and garbage liner items are likely not equivalent. These data also show that there was a large difference in performance between the two groups on the garbage liner item, which may be due to the sample size of the groups. Future testing with more items and larger sample sizes should be done to create and identify equivalent items.

As with the problem-finding items, the item correlation may be improved with more equivalent items. It could also be that there are other confounding factors at work in these measurements. For example, if a person's past experience had led them to deal with one of these problems before, they may already have solutions in mind for these problems. Future researchers may need to look for problems to use as prompts that are either universally familiar or universally unfamiliar to the population that is being tested.

**Results for Communication Items**

For the revised test, the communication items were changed to go with the solution items rather than the ranking items. The instructions were also changed to be clearer and describe what the raters were looking for in the items. Table 10 contains the resulting data.

**Table 10**
*Summary of Statistics for Communication Items*

| | Overall | | Group C | | Group D | | Item correlation |
|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | |
| Headphone pitch | 8.62 | 1.41 | 9.10 | 1.37 | 8.11 | 1.25 | 0.43 |
| Garbage liner pitch | 8.28 | 1.28 | 8.20 | 1.50 | 8.37 | 0.98 | |

These data show that even though the communication items use the same wording, they are not necessarily equivalent. The difference between the scores was more pronounced in Group C than in Group D. It is not clear why this happened, but it could be that a larger data set is needed to stabilize the results. There may be some statistical anomaly in one of the groups that would disappear with a larger test sample. Some of the differences may come from the differences in the problem statements from the solution items. More testing would need to be done with different prompts in the solution items. It may be found that solution items with more equivalence could lead to communication items with more equivalence also. Because the communication items rely so heavily on the solution items, the lack of correlation for the solution items is likely contributing to the lack of correlation for the communication items. In future studies, researchers should see how the item correlations for the communication items change as the item correlations for the solution items improve.

Interrater reliability for the revised test was also high. The correlation between the raters' scores on the two items were 0.76 and 0.74, respectively. This is enough to confidently claim "good" interrater reliability (Cicchetti, 1994).

**Results for Ranking Items**

The ranking items were chosen from the items in the first round of testing. The bike seat and toilet items were chosen for the revised test because they were the higher scoring items from the previous test. Table 11 shows the summary statistics.

**Table 11**
*Summary of Statistics for Ranking Items*

| | Overall | | Group C | | Group D | | Item correlation |
|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | |
| Bike seat | 4.64 | 2.90 | 4.15 | 2.85 | 5.16 | 2.85 | 0.09 |
| Toilet | 7.21 | 2.40 | 7.30 | 2.22 | 7.11 | 2.57 | |

The data in the table show that the order effect and fatigue problems were reduced but that the difference in the item difficulties became more pronounced. Both groups performed better on the toilet item than on the bike seat item.

The item correlation for these items was very low, indicating that there is a serious problem with these items. The problem likely comes from the lack of agreement between expert rankings. With more consensus in the expert rankings, it is likely that the item correlations will improve because there will be a stronger standard against which students can be compared.

## Conclusion

The Innovation Test Instrument (ITI) was create to address the need for an innovation test that assesses an individual's ability to perform all of the different parts of the process of innovation (Lewis, 2011). The purpose of this article was to outline the design, development, implementation, and validation of the ITI, which was designed to test an individual's innovative capacity in the skills identified from the literature: idea finding, idea shaping, idea defining, idea refining, and idea communicating. The findings from this study helped the researchers to improve the test and argue for initial validity based on the high reliability from interrater scores. Nonetheless, a more in-depth validation study of ITI would be valued. Below, the issues of validity and reliability are discussed briefly.

### Validity

Although more testing should be done to further establish validity of the scores from this instrument, this study showed that there is a good case for some types of validity-related evidence: content-related evidence, consequence-related evidence, construct-related evidence, face validity evidence, and criterion-related evidence of validity.

Content-related evidence is the degree to which an instrument covers the content within a specific domain (Babbie, 1990). The evidence criterion is fulfilled by the description of the processes of innovation as outlined in this paper, and used to design the instrument (as described above). In addition, the method of development and implementation of the ITI also helped to establish a link between the instrument and the content that is to be tested. The review of literature showed that the BYU Innovation Bootcamp curriculum is aligned with other innovation processes and models, and the methods employed shows that the ITI is aligned well with the Bootcamp curriculum.

According to Miller (2009) consequence validity describes the thoughtfulness of the consequences of use and interpretation of assessment results. In this study, the stakes of the test results were very low. Results were not used to establish grades for students or determine whether they should be admitted to certain programs or positions. The only real consequence of the results of this instrument in its current form is that the results could affect how the Innovation Bootcamp is taught in the future. The results of this instrument should not be used for other considerations without further study.

In this article, the development of the test items was described, showing that the test items were developed using generally accepted test development practices. This can be a positive initial step in establishing construct-related evidence of validity. Construct validity refers to how well the measurements taken in an assessment relate to each other according to theoretical constructs (Babbie, 1990). Showing that appropriate methods were used does not establish construct validity on its own, but it does show that construct validity is more likely than if they had not been used.

Construct-related evidence was also addressed in the revisions that were made between the two rounds of testing. Changing the pictures in the problem-finding items, moving the communication items, revising the communication items' instructions, shortening the instrument, and adding incentives were all ways that the researchers reduced construct-irrelevant variance.

Face validity is a type of validity that refers to how much the respondents perceive that the test is relevant or important (Miller et al., 2009). The first round of testing showed that the instrument had some face validity for the students of the Innovation Bootcamp. Even though test fatigue caused results that made some interpretations difficult, the fact that so many students participated as much as they did demonstrates a level of face validity. This improved more in the second round of testing because students were more invested in completing the test well. Some students commented that they enjoyed taking the test or thought that it was an interesting way to practice what they had learned in the Innovation Bootcamp. The fact that students felt that the test was relevant to what they had learned is a strong piece of evidence in favor of face validity.

Criterion-related evidence refers to how well a measured variable can predict other variables. In this test, a claim of criterion validity would say that scores on this test are a good predictor of how likely a person is to be a strong innovator. This type of validity was not formally studied in this research. Notwithstanding, the researchers of this study made anecdotal observations that support criterion validity. The researchers of this research also assisted in the instruction of the Innovation Bootcamp. The researchers noted that the top scorers on the test were also students who had many innovative ideas during the Innovation Bootcamp. This alone is not enough to establish criterion validity, but it's an initial value to be considered.

**Reliability**

In this study, two types of reliability were studied: test form equivalence and interrater reliability. The results discussed in detail the equivalence of the items. Because of the differences in the means scores of the items, all of the item types in this instrument need additional work before they can be used for pre–post testing of the Innovation Bootcamp. Even though this instrument did not achieve form equivalence, it is a strong first attempt that will facilitate future instrument development in the area of innovation assessment.

Although the means and standard deviations for the items show that these items are not equivalent, they can still be used as pre- and post-test items to measure the impact of the Innovation Bootcamp. This can be done by using the data from this sample to compute *z*-scores for the responses to each item. For example, in this study, the garage item had a mean of 13.00 and a standard deviation of 6.14, and the bedroom item had a mean of 9.69 and a standard deviation of 5.89. If a student did the garage item in a pretest and scored 11, the *z*-score (in relation to the sample group from this study) would be -0.33. If the student did the bedroom item as part of a posttest, and scored 10, the *z*-score would be +0.05. In this case, the positive change in the *z*-score would show that the student performed better on the posttest item than on the pretest item.

The interrater reliability for the communication items was also tested. In the first round of testing, interrater reliability levels were 0.94 and 0.97, and in the second round, interrater reliability levels were 0.76 and 0.74. According to Cicchetti (1994), interrater reliability between .60 and .74 is considered "good." This leads the researchersto be confident in the interrater reliability of the scores for the communication items.

## References

Anderson, L. W., & Krathwohl, D. R. (Eds.) (with Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C.). (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abr. ed.). New York, NY: Longman.

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.

Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail.* Boston MA: Harvard Business School Press.

Christensen, C. M., & Eyring, H. J. (2011). *The innovative university: Changing the DNA of higher education from the inside out*. San Francisco, CA: Jossey-Bass.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Drucker, P. F. (1985). *Innovation and entrepreneurship: Practice and principles*. New York, NY: Harper & Row.

Dyer, J., Gregersen, H., & Christensen, C. M. (2011). *The innovator's DNA: Mastering the five skills of disruptive innovators*. Boston, MA: Harvard Business School Press.

European Commission. (2011, December 12). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Open data: An engine for innovation, growth and transparent governance*. Brussels, Belgium: Author. Retrieved from http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2011/0882/COM_COM%282011%290882_EN.pdf

Fagerberg, J. (1999). The need for innovation-based growth in Europe. *Challenge*, *42*(5), 63–79. doi:10.1080/05775132.1999.11472122

Friedman, T. L., & Mandelbaum, M. (2011). *That used to be us: How America fell behind in the world it invented and how we can come back.* New York, NY: Farrar, Straus and Giroux.

Getzels, J. W. (1975). Problem-finding and the inventiveness of solutions. *Journal of Creative Behavior*, *9*(1), 12–18. doi:10.1002/j.2162-6057.1975.tb00552.x

Hasso Plattner Institute of Design at Standford University. (2017). Bootcamp bootleg. Retrieved from https://dschool.stanford.edu/resources/the-bootcamp-bootleg?rq=Bootcamp%20Bootleg

Houtz & Krug (1995). *Assessment of creativity: Resolving a mid-life crisis*. Educational Psychology Review, 7, 269-300.

Howell, B., Skaggs, P., & Fry, R. (2010). The innovation boot camp. In W. Boks, C. McMahon,W. Ion, & B. Parkinson (Eds), *Proceedings of the 12th International Conference on Engineering and Product Design Education* (pp. 216–221). Backwell, Bristol, United Kingdom: The Design Society. Retrieved from https://www.designsociety.org/publication/30129/the_innovation_boot_camp

IDEO. (2017). About IDEO. Retrieved from http://www.ideo.com/about/

Innosight. (2011). Our approach. Retrieved from http://www.innosight.com/our_approach/create_or_reshape_process.html?gclid=CIPCytTUsKgCFSUZQgod0BZJHA

Kelly, T. (with Littman, J.). (2005). *The ten faces of innovation: IDEO's strategies for defeating the devil's advocate and driving creativity throughout your organization*. New York, NY: Currency/Doubleday.

Lewis, T. (2011). *Creativity and innovation: A comparative analysis of assessment measures for the domains of technology, engineering, and business* (Master's thesis, Brigham Young University). Retrieved from http://scholarsarchive.byu.edu/etd/2865/

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). Measurement and assessment in teaching (10th ed.). Upper Saddle River, NJ: Pearson.

National Portal of India. (2014). Innovation and business: Importance or benefits. Retrieved from http://www.archive.india.gov.in/business/innovation/benefits.php

Radosevic & Mickiewicz (2003). Innovation Capabilities in Seven Candidate Countries: An Assessment, Vol. 2.8. Brussels: Enterprise Directorate General, CEC, 2003.

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). New York, NY: Free Press.

Stanford Graduate School of Business. (2017). Design thinking boot camp: From insights to innovation. Retrieved from https://www.gsb.stanford.edu/exec-ed/programs/design-thinking-bootcamp

Torrance, E. P. (1963). *Creativity*. Washington, DC: American Educational Research Association & National Education Association.

Wagner, T. (2010). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need—And what we can do about it*. New York, NY: Basic Books.

Wagner, T. (2012). *Creating innovators: The making of young people who will change the world* (Rev. ed.). New York, NY: Scribner.

The White House: President Barack Obama. (2011, January 25). Remarks by the President in State of Union Address. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2011/01/25/remarks-president-state-union-address

## About the Authors

**Jacob Wheadon** (jacob.wheadon@gmail.com) is a graduate student of Technology and Engineering Studies at Brigham Young University.
**Geoff A. Wright** (geoffwright@byu.edu) is Associate Professor of Technology and Engineering Studies at Brigham Young University.
**Richard E. West** (rickwest@byu.edu) is Associate Professor of Instructional Psychology and Technology at Brigham Young University.
**Paul Skaggs** (paul_skaggs@byu.edu) is Associate Professor of Industrial Design at Brigham Young University.

## Scope of the JTE

The *Journal of Technology Education* provides a forum for scholarly discussion on topics relating to technology and engineering-related education. Manuscripts should focus on technology and engineering-related education research, philosophy, and theory. In addition, the *Journal* publishes book reviews, editorials, guest articles, comprehensive literature reviews, and reactions to previously published articles.

Technology and Engineering Education (T&EE) is a program that resides at the P-12 school levels for all students and at post-secondary institutions for those students interested in teaching or obtaining employment in the technology or engineering fields. Technology and engineering education is primarily taught by technology and engineering teachers, with a focus on engineering design. T&EE may be considered a stand-alone discipline or part of a larger discipline in science, technology, engineering, and mathematics (STEM). Regardless of the approach, T&EE focuses on technological literacy and engineering design; engineering design is the verb tense of engineering.

At the P-12 grade levels, the goal is for students to develop technological and engineering literacy, regardless of career aspirations, through hands-on, contextual applications of technological and engineering concepts. T&EE students, use a hands-on approach to solve technological problems using problem solving and creativity, while working under constraints, which involves the use of optimization and predictive analysis. At the P-5 grade levels, technology and engineering concepts are integrated into existing coursework such as reading, mathematics, science, and social studies. Typical courses students would take at the 6-12 grade levels in a T&EE program would consist of (a) information and communication technologies, including computer-aided drafting and design, (b) engineering design, (c) construction technology, (d) manufacturing technology, (e) energy, power, and transportation technology, and (f) medical, agricultural, and related biotechnologies. Within these courses, students would utilize troubleshooting, research and development, invention and innovation, and problem solving. The focus of T&EE at the P-12 levels is not to prepare future engineering majors/students, but to provide an education for all students.

## Editorial/Review Process

Manuscripts that appear in the *Articles* section have been subjected to a blind review by three or more members of the Editorial Board. This process generally takes from six to eight weeks, at which time authors are notified of the status of their manuscript. Book reviews, editorials, and reactions are reviewed by the Editor.

**Manuscript Submission Guidelines**

One paper copy of each manuscript and an electronic version in Microsoft Word format should be submitted to:

Chris Merrill, JTE Editor
Department of Technology
Illinois State University
215 Turner Hall
Normal, IL 61790-5100

1. Overseas submissions in Microsoft Word format may be sent electronically via the Internet (to cpmerri@ilstu.edu) to expedite the review process.
2. All manuscripts must be double-spaced and must adhere to the guidelines published in *Publication Guidelines of the American Psychological Association* (6th Edition). **Tables and figures, however, should be imbedded within the text itself rather than at the end of the document.**
3. **All figures and artwork must be scalable to fit within the JTE page size (4.5” x 7.25” column width and length) and included electronically within the document.**
4. Line drawings and graphs must be editable within Microsoft products and in vector rather than raster format when possible.
5. **Shading should not be used as a background for illustrations or graphs and within bar graphs.** If needed, fill patterns consisting of lines should be used.
6. Manuscripts for articles should generally be 15-20 pages (22,000-36,000 characters in length, with 36,000 characters an absolute maximum). Book reviews, editorials, and reactions should be approximately four to eight manuscript pages (approx. 6,000-12,000 characters).
7. Authors for whom English is not the primary language must enlist a native English editor for the manuscript prior to submission. This person and his/her email address must be identified on the title page of the manuscript.

*Subscription Information*

The *Journal of Technology Education* is published twice annually (Fall and Spring issues). New and renewing subscribers should copy and mail the form below:

Name (please print) ______________________________________________

Mailing Address (please print) ____________________________________

_________________________________________________________________

Email address: ____________________________Fax: _______________

☐New Subscription ☐ Renewal Subscription

Make checks payable to: *Journal of Technology Education*. All checks *must* be drawn on a US bank.

Regular (USA): $20
Regular (Canada/Overseas): $30
Library (USA): $30
Library (Canada/Overseas): $40
Individual Back Issues (USA): $10 each
Individual Back Issues (Canada/Overseas): $15 each

Return remittance along with this form to:

Chris Merrill, JTE Editor
Department of Technology
Illinois State University
215 Turner Hall
Normal, IL 61790-5100

*JTE Co-Sponsors & Membership Information*

The International Technology and Engineering Educators Association (ITEEA) is a non-profit educational association concerned with advancing technological literacy. The Association functions at many levels – from international to local – in responding to member concerns. The Council on Technology and Engineering Teacher Education (CTETE), affiliated with the ITEEA, is concerned primarily with technology teacher education issues and activities. For membership information, contact: ITEEA, 1914 Association Drive, Reston, VA 22091 (703) 860-2100.

*Electronic Access to the JTE*
*All* issues of the *Journal of Technology Education* may be accessed on the World Wide Web at: http://scholar.lib.vt.edu/ejournals/JTE/ (Note: this URL *is* case sensitive).

Directory of Open Access Journals Statement
As an open access journal, the *JTE* does not charge fees for authors to publish or readers to access.