Measuring Student Content Knowledge, iSTEM, Self Efficacy, and Engagement through a Long-Term Engineering Design Intervention

Dr. Michael A. de Miranda, Colorado State University

Professor Engineering Education, School of Education and Department of Electrical and Computer Engineering. Colorado State University, Fort Collins, CO USA

Dr. Karen E. Rambo-Hernandez, West Virginia University

Karen E. Rambo-Hernandez is an assistant professor at West Virginia University in the College of Education and Human Services in the department of Learning Sciences and Human Development. In her research, she is interested the assessment of student learning, particularly the assessment of academic growth, and evaluating the impact of curricular change.

Dr. Paul R. Hernandez, West Virginia University
Measuring Student Content Knowledge, iSTEM, Self Efficacy, and Engagement Through a Long Term Engineering Design Intervention

Abstract
The current study reports on the outcomes of a classroom-based long-term engineering design intervention intended to increase high school students’ perceptions of the integrated nature of STEM disciplines (iSTEM) and to assess the effect of the intervention on student participation in an extracurricular STEM activity (i.e., a research poster symposium). Cross-disciplinary teams of students (n=373) from high school mathematics, science, and engineering classrooms completed engineering design challenges. The results indicated that, consistent with our predictions, the intervention exhibited a positive impact on students that began the study with the lowest iSTEM scores. Furthermore, the classroom environment mattered. While no individual scores (i.e., posttest iSTEM scores) were predictive of participation in the poster symposium, the collective scores were (i.e., mean classroom iSTEM scores).

Four measures were used in this study; Content knowledge quiz. Student content knowledge was assessed with a teacher made nine-item multiple-choice quiz; self-efficacy, task specific self-efficacy was assessed through a nine item measure; iSTEM perceptions. Participants responded to a nine-item iSTEM scale developed and validated by the authors in a previous study, to measure student perceptions of the interconnections between mathematics, science, and engineering; and STEM clubs. Participants responded “Yes” (1) or “No” (0) to the question regarding their involvement in extracurricular STEM club.

Hierarchical linear modeling (HLM) was used in this analysis because it distinguishes variability in scores at the student-level (i.e., level-1) from variability in scores at the classroom level (i.e., level-2), which results in correctly estimating standard error. Therefore, HLM was used to conduct multilevel-paired sample t-tests. Further, all analyses were conducted with Restricted Maximum Likelihood estimation. The results indicated that, consistent with our predictions, the intervention exhibited a positive impact on students that began the study with the lowest iSTEM scores. Furthermore, the classroom environment mattered. While no individual scores (i.e., posttest iSTEM scores) were predictive of participation in the poster symposium, the collective scores were (i.e., mean classroom iSTEM scores).

Introduction
Many science, technology, engineering and mathematics (STEM) concepts, especially those learned in the critical formative years of pre-collegiate education are abstract in nature, often taught in vertically articulated course offerings that are frequently unconnected horizontally with other STEM course content. The lack of concept and content connections to authentic applications makes learning difficult for young learners. In addition, few opportunities exist within K-12 education for students to apply STEM learning in contextually authentic learning-in-doing inquiry and design driven environments in which they are immersed over time greater than a few class periods. Combine these factors with student misconceptions of what engineering
practice is and less than optimal instructional models yields a volatile combination for student attrition and low perceived value for learning STEM subjects.¹

The aversion to learning basic STEM concepts due to their high abstractness, low perceived value, utility, and disconnection from applications has triggered a decrease in confidence in STEM learning among entering college students. This can be illustrated by the fact that enrollment in U.S. institutions of higher education has grown steadily at all levels rising from 14.5 million students in 1994 to 20.7 million in 2009, but such a growth is not fully reflected in science and engineering. Institutions of higher education in the United States granted engineering degrees in the mid-2000s at a lower rate than in the mid-1980s. The number of American students earning bachelor’s degrees increased by 16% over the past 10 years, however, the number of bachelor’s degrees earned in engineering decreased by 15%. Nationally, less than 50% of the students who enrolled in engineering curriculum complete the program. American Universities typically lose 50% of engineering freshmen and sophomore during the first two years of their engineering program. This trend is continuing in the foreseeable future and it can be attributed to (at least) several factors:

- The traditional teaching of math, physics, and engineering concepts are isolated. Each discipline operates within its own silo. Students do not see the relationship of what is taught to what they are interested in learning.²
- Early engineering students fail to identify with and become part of the engineering community through practice, inclusion, and engagement.³
- Only small populations of high school students find themselves attracted to engineering schools and have never experienced doing research or engineering design.⁴

Addressing these significant factors in the learning of STEM and especially in coming to know, experience, and integrate engineering practices as part of the STEM learning continuum is becoming an imperative that pre-collegiate education must address. However, challenges exist when a shift in paradigmatic approach to learning and instruction is introduced to a well-established educational system.

**Shifting approaches to STEM education**

The recent release of the Next Generation Science Standards (NGSS) marks a significant shift in the core concepts and approaches guiding science, technology, engineering, and mathematics education content in the coming years.⁵ Most notable is the inclusion of engineering and technology concepts in a framework that emphasizes practices, crosscutting concepts, and core ideas. The repositioning of engineering and technology content within science education brings to light new opportunities and challenges when conceptualizing the design and delivery of instruction in STEM subjects. Moreover, realizing the full potential of the NGSS will require new conceptions of learning and instruction being adopted to include the richness of unifying practice, inquiry, and design across STEM concepts and contexts.

The NGSS articulates a broad set of expectations for students in science grounded in practices and inquiry. Within these guiding standards are three major dimensions around which grades K-12 science education needs to be integrated into standards, curriculum, instruction, and assessment. These dimensions include: scientific and engineering practices; crosscutting concepts that unify the study of science and engineering through their common application
across fields; and core ideas in four disciplinary areas: physical sciences, life sciences, earth and space sciences, and engineering, technology, and applications of science.\textsuperscript{5}

Integrating the three dimensions of scientific and engineering practice, crosscutting concepts, and disciplinary core ideas that cover traditional scientific fields of study (i.e. physical science, life science, and earth and space science) now includes the addition of engineering, technology, and applications of science. Integrating the three dimensions could prove illusive, however approaches informed by research on teaching and learning from cognitive sciences combined with aggressive methodological approaches to measuring student learning within the three dimensions can yield promising results. John Bruer\textsuperscript{6} in his seminal book \textit{Schools for Thought} argued that, ‘the National Assessment of Educational Progress (NAPE; often referred to as the Nations report card) results indicate that current curricula, teaching methods and instructional materials successfully impart facts and rote skills to most students but fail to impart high-order reasoning and learning skills’ (p. 5). This statement continues to resonate today as it did in 1993. Other researchers have explored transforming the classroom from “work sites where students perform assigned tasks under management of teachers into communities of learning and interpretation, where students are given significant opportunity to take charge of their own learning…attempting to engineer an innovative educational environment”\textsuperscript{7} p.141

Grounding of the intervention design

**STEM content knowledge.** The conceptualization and design of this study is informed by two perspectives; the first influenced from well researched areas of teaching and learning from the cognitive sciences; and second from the newly released NGSS.\textsuperscript{5} The consonance between models of classroom learning and teaching informed by research from the cognitive sciences and the new frameworks vision to actively engage students in scientific and engineering practices and apply crosscutting concepts to deepen their understanding of the core ideas in these fields. In addition, the new vision sets a goal of students gaining sufficient knowledge of the practices, crosscutting concepts, and core ideas of science and engineering to engage in public discussions about these subjects and communicate their understanding of science and engineering practices.\textsuperscript{5}

One approach to this transformation process can be addressed through adopting well-researched approaches to learning and instruction grounded in the cognitive sciences. An example of this has been evidenced with the emergence of engineering and technology education as an integral component of STEM education. For many years technology education alone struggled to establish itself as an equal partner in general education and often was unable to gain recognition for the value of its instruction. Often technology educators touted the effectiveness of their hands-on “making” programs based on anecdotal evidence gathered from their classroom experiences reflecting how their instructional methods empower students to learn. Today’s engineering and technology education originated without any meaningful input from cognitive science research. However, it appears that engineering and technology education practices advocated in the new K-12 science education framework are remarkably consonant with findings from cognitive science that defines good instruction.\textsuperscript{8}

The foundations of cognitively based models of instruction hold four elements of learning common across the various instructional strategies:\textsuperscript{8}

1. Students learn to engage actively with the learning process and content.
2. Through the instructional design, students learn to reflect and use existing structures of knowledge to guide and further their learning.

3. Students learn to interact in classrooms or communities of learning where knowledge and information are shared openly in an environment that values participation and interaction between students, teachers, and sources of knowledge outside the classroom.

4. The engineering component of learning and instruction emphasizes the process and design of solutions instead of the solutions themselves.

This approach allows students to explore mathematics and science in more personalized context while helping them to develop the critical thinking and reflection skills that can be applied to all facets of their work and academic lives. The American Society of Engineering Education (ASEE) has continually promoted the notion that, engineering design, by its very nature, is a pedagogical strategy that promotes learning across disciplines. In addition, when instruction is organized to promote cross-disciplinary interaction, critical debate, design solution finding and problem focus without classroom or school boundaries, the ethos of the learning environment is fundamentally changed from knowledge transmission sites to sites of discourse, knowledge sharing and coming to know and learn how knowledge is applied and shared.

**STEM self-efficacy.** Social cognitive career theory (SCCT) helps to explain why student choose and persist into careers, particularly those in STEM fields. Self-efficacy, the belief held by students about their ability to perform or complete a task, is one of the cornerstones of SCCT. Consistently, self-efficacy has been predictive of career choices, persistence toward a career, and performance. Self-efficacy should be assessed as it directly relates to a specific task or skill not just a broad conceptualization of ability, and while related to cognitive ability, a student’s task specific self-efficacy is uniquely related to career related decisions. Further, within SCCT, self-efficacy is not static but is rather dynamic and sensitive to interactions with others and with social factors that students encounter. Thus, studies that incorporate opportunities for students to engage with new material, such as engineering concepts, within a social setting should expect to see increases in student task-specific self-efficacy. Given the outcomes modeled in SCCT, increases in engineering task-specific self-efficacy should translate to a greater likelihood to choose, persist, and eventually become engineers.

**STEM associational fluency.** The integrated approach to STEM education (iSTEM) includes instructional approaches and complex classroom interventions that interweave content and learning experiences among and between any of the STEM subjects or other school subjects. While learning-in-doing through integrated design problems situates the learner to apply the “thinking tools” from varied STEM knowledge structures, e.g. mathematics or physics principles on demand when needed. This STEM “associational fluency”, where students thinking crosses content borders begins to mirror expert engineering practice. Prior research with students indicates that they often view science, technology, engineering, and mathematics as separate fields, and thus perceive relatively low levels of integration across STEM disciplines. However, research among high school students indicates that engaging in classroom based projects that vertically and horizontally integrate applications of the math, science, and engineering improve students’ appreciation of the integrated nature of STEM disciplines. Furthermore, this research also shows that engaging student in projects that integrate across
STEM disciplines is impactful on those that have relatively low perceptions of the integrated nature of STEM.

Research questions
RQ1: Do students make gains in content (e.g., physics) self-efficacy or knowledge (i.e., quiz scores) over the course of the intervention?
RQ2: Does that pattern of gains in iSTEM scores match the pattern described in prior research (i.e., students with initially low iSTEM scores exhibit the most gains in iSTEM perceptions)?

Method
Participants
The four schools had differing groups of classes that worked together to complete the projects: (a) School 1 (Anatomy & physiology; engineering and technology; geometry); (b) School 2 (engineering and technology; geometry; general physics); (c) School 3 (biology; statistics; engineering and technology); and (d) School 4 (calculus; general and AP physics; engineering and technology). The total number of participating students across the four schools was 373. Of the overall sample, 73 participants were missing data at either pretest or posttest and were thus excluded from the analytic sample. The final analytic sample for this study consisted of high school students in the 9th (20%), 10th (39%), 11th (19%), and 12th (22%) grades. The sample was nearly evenly split by gender (54% male & 46% female). Further, the sample was primarily White (83%), with a smaller proportion of students being Hispanic/Latino (7%), Asian (4%), multi-ethnic (3%), or Other (3%). Finally, only a small proportion of students (16%) were involved in extra-curricular STEM clubs (e.g., robotics club, web design club, 4-H).

Procedure and intervention organization
This study examines the impact of a complex classroom intervention that addresses several “game changing” factors that could influence pre-collegiate STEM learning.
1. Mathematics, science, and engineering and technology teachers worked collaboratively to blueprint their individual curriculum and built content matrices that identified content intersections that would reinforce the applications of content in authentic practice.
2. Students worked in cross-functional design teams made up of students from each of the three content classrooms.
3. The intervention extended past the traditional several week-long unit to steep students in design thinking, decision making, design of experiments, testing, data collection and analysis, optimization, and communication of final design solutions with a team engineering and math/science research posters presented publically.
Teacher teams consisting of math, science, and engineering and technology teachers from four high schools gathered during the summer prior to the intervention to co-plan and blueprint the subject matter content each teacher was planning to cover during the semester the long-term engineering design intervention was to occur. The teams were lead through the content blueprinting process by the research team. The blueprinting process made explicit the specific learning concepts, defined or operationalized the content, identified the level of knowledge learning desired according to Blooms Taxonomy, identified common student misconceptions related to the concept being learned, and identified the test item number associated with measuring the content as shown in appendix A. Teachers then created and shared the
assessments used to measure student learning as shown in appendix B. Subject matter experts using the protocol shown in appendix C reviewed each assessment item for;

1. Construct Rating: Reviewer associates item with construct identified in content blueprint.
2. Confidence Rating: Reviewer rates HOW CERTAIN they are that the item fits into the content construct.
3. Relevance Rating: Reviewer rates how RELEVANT/APPROPRIATE the item is to the construct (i.e., how representative of the construct is the item).
4. Reviewer double checks the correct answer identified in the answer key.
5. Reviewer identifies any distracters that do not align with “misconceptions” (listed in the blueprint) or are not relevant to the construct.

Following the content test review teachers were provided with feedback to revise and edit the items. Nine of the best items representative of the major concepts being taught were selected for the content assessment.

The design problem intervention
Impactful pre-collegiate engineering education must reflect what engineers do. Therefore, the complex classroom intervention used in this study translated the activities from an active interdisciplinary biomedical engineering research program. The research consisted of activities in silicon nano scale sensor design, modeling, and understanding how molecules move and the functions of multi-cellular tissues and organ systems in response to external chemical and physical stimuli through intercellular communication. The research is critical for continued understanding and advances in some of the fundamental questions facing biology and medicine, and ultimately our society for better quality of life. The research and design of this biosensor require the cooperation of professors and graduate students in biology, chemistry, computer science, electrical and computer engineering, and mathematics.

To mirror the practice of this interdisciplinary research students participating in this study were challenged to design and test “sensing” related problems of their choice. For example, teams made up of math, anatomy/physiology, and engineering and technology students designed bicycle helmets fitted with sensors to test impact absorption and collect data related to helmet materials and design to reduce force transfer to the head as shown in figure 1. Students formed interdisciplinary teams of engineering, math, and science students along with their STEM content teachers. Students were required to conduct research, design experiments and engineer test beds that enabled them to execute their experiments and collect data.

Figure 1. Helmet sensor design team.

Early in the fall semester (2013), prior to the intervention implementation, teachers scheduled with the research team to administer pretests. All involved were required to adhere to the guidelines required under human subjects’ approval (Office of Regulatory Compliance) and complete an informed assent to participate in research form. To protect the student identities and ensure anonymity, teachers assigned their students unique research identification numbers and created a classroom link list. Research team members then collected the informed assessment
forms and administered the pretest. Following pretest, the intervention was implemented in each classroom. The intervention (i.e., engineering design challenge) proceeded for 16 to 24 weeks (varying by class). After the completion of the intervention, teachers schedule with the research team to administer posttests. Finally, and concurrent with posttests, individual students in the classes joined into teams to develop and present a final research or engineering posters at a symposium held on a university campus.

**Measures**

**Content knowledge quiz.** Student content knowledge was assessed with a teacher made nine-item multiple-choice quiz. As described above, the content quiz consisted of multiple-choice items covering the major concepts covered in the class and as part of the design project. For example, students in one engineering design class responded to items covering applications of Ohm’s Law serial and parallel circuits. Each quiz item contained a single correct response and three distracters (i.e., incorrect but plausible response options). Students scale scores on the content knowledge quiz were produced by taking the proportion of correct answers at pretest and posttest. Therefore, individual scores could range from 0 (no correct answers) to 1 (all correct answers).

**Self-efficacy.** Task specific self-efficacy was assessed through a nine item measure. Specifically, after each multiple-choice question on the quiz students responded to the question “How confident are you in your answer to question X?” (where “X” was 1, 2, 3, …, 9). The response options were scored from 1 (“Not at all confident”) to 5 (“Completely confident”). Students scale scores on the self-efficacy instrument were produced by taking the mean across responses. Therefore, individual scores could range from 1 to 5, with higher scores indicating higher self-efficacy, see Table 1 for descriptive statistics.

**iSTEM perceptions.** Participants responded to a nine-item scale developed and validated in a previously published study, to measure student perceptions of the interconnections between mathematics, science, and engineering. In validating the iSTEM scale in a previous study similar to the one reported here, a single factor Confirmatory Factor Analysis (CFA) model was fit to the data based on the model identified through Exploratory Factor Analysis (EFA) at pretest. The CFA analysis was conducted using maximum likelihood estimation in AMOS version 18. The analysis of the single factor model revealed excellent model fit ($\chi^2 (df = 20) = 29.11, p = .08$; comparative fit index = .98; and root mean square error of approximation [RMSEA] = .05, 90% CI [.00, .09]). Further, an inspection of the factor model revealed that all items exhibited high factor loadings, see Figure 2. A reliability analysis at of posttest scores indicated highly similar findings to those found at pretest (i.e., Cronbach’s alpha = .85, $N = 170$, $M = 3.50$, $SD = 0.69$, KS $z = .05$, $p = .20$).
Figure 2. CFA factor structure of STEM Connections scale
Note: Q9 had been set as the marker variable. Factor loadings values in the figure are standardized.

Students responded to statements, such as “I have applied connections between mathematics, science, and engineering to help me solve problems outside of school”, on a five-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). Students scale scores on the iSTEM instrument were produced by taking the mean response across items. Therefore, individual scores could range from 1 to 5, with higher scores indicating higher iSTEM perceptions, the descriptive statistics for this study is shown in table 1 in the results section.

STEM clubs. Participants responded “Yes” (1) or “No” (0) to the question regarding their involvement in extracurricular STEM clubs: “Do you participate in any Math, Science, Engineering, or Technology clubs inside or outside of school?” If the student indicated “Yes,” s/he was asked to specify the name of the STEM club, see descriptive statistics in table 1 in results section.

Descriptive statistics
Prior to testing hypotheses, we examined patterns and associations. At the student level, consistent with expectations participants’ self-efficacy and quiz scores showed positive gains from pretest to posttest, see table 1. Further, consistent with prior research and our expectation, iSTEM scores show no mean level gain from pretest to posttest. Regarding associations, efficacy, quiz, and iSTEM scores exhibited moderately positive correlations (i.e., moderate stability) from pre- to posttest. Further, consistent with theory, self-efficacy and quiz scores were moderately and positively correlated at both pre- and posttest.

Results
Table 1.
Intercorrelations and Descriptive Statistics of Demographic Characteristics, Self-Efficacy, Quiz, and iSTEM Scores at Pre- and Post-test (N = 300)

<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Self-Efficacy (Pretest)</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Quiz (Pretest)</td>
<td></td>
<td>.63***</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. iSTEM (Pretest)</td>
<td></td>
<td>.06</td>
<td>.06</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Self-Efficacy (Posttest)</td>
<td>.60***</td>
<td>.40**</td>
<td>.16**</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Quiz (Posttest)</td>
<td></td>
<td>.48***</td>
<td>.62***</td>
<td>.10</td>
<td>.57***</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. iSTEM (Posttest)</td>
<td></td>
<td>.04</td>
<td>.08</td>
<td>.71***</td>
<td>.26***</td>
<td>.17**</td>
<td>.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Female (0 = male, 1 = female)</td>
<td>- .12*</td>
<td>.07</td>
<td>-.17**</td>
<td>-.07</td>
<td>.06</td>
<td>-.22***</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. White</td>
<td></td>
<td>-.01</td>
<td>.05</td>
<td>-.06</td>
<td>.03</td>
<td>.05</td>
<td>.04</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>(0 = non-White, 1 = White non-Hispanic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. STEM Clubs</td>
<td></td>
<td>.14*</td>
<td>.03</td>
<td>.31***</td>
<td>.15**</td>
<td>.02</td>
<td>.25***</td>
<td>.00</td>
<td>.07</td>
</tr>
<tr>
<td>(0 = did not participate, 1 = participated in STEM clubs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>3.07</td>
<td>0.55</td>
<td>3.83</td>
<td>3.71</td>
<td>0.64</td>
<td>3.82</td>
<td>0.46</td>
<td>0.83</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.97</td>
<td>0.25</td>
<td>0.64</td>
<td>0.67</td>
<td>0.24</td>
<td>0.68</td>
<td>0.50</td>
<td>0.37</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note: Cronbach’s alpha coefficient reported in italics on the diagonal.

* p < .05, ** p < .01, *** p < .001
Self-efficacy and quiz scores
The main research questions in this study concerned the change in student knowledge, self-efficacy, and iSTEM perceptions through their involvement in an engineering design challenge (i.e., intervention). The most direct method of assessing such changes was a paired samples t-test (i.e., posttest score – pretest score) to test for statistically significant gain in scores. Unfortunately, the design of this study (i.e., the students were nested within classrooms) made it inappropriate to use a standard paired samples t-test, as such a test would underestimate the standard error of the statistical test. Hierarchical linear modeling (HLM) is a more appropriate statistical method in nested designs because it distinguishes variability in scores at the student-level (i.e., level-1) from variability in scores at the classroom level (i.e., level-2), which results in correctly estimating standard error. Therefore, HLM v. 7.01 was used to conduct multilevel paired sample t-tests. Further, all analyses were conducted with Restricted Maximum Likelihood estimation.

More specifically, first, we derived gain scores (i.e., posttest score – pretest score) for each construct (i.e., quiz, efficacy, & iSTEM). Second, we entered the gain scores into HLMs as the dependent variable.

To formally test research question-1, a null model (i.e., intercept only with no predictors) was fit to self-efficacy the gain scores. An examination of the fixed effects indicated that the intercept (i.e., mean gain) was positive and statistically significant, see table 2 (Self-Efficacy Gain). The model revealed that self-efficacy scores gained 0.64 points from pre- to posttest. Finally, an examination of the random effects revealed that an extremely large proportion of the variability in gain scores was between classrooms. Specifically, the intraclass correlation coefficient was .48, which indicated that 48% of the variability in gain scores was between classrooms. A further analysis, regressing self-efficacy gain scores on the control variables revealed that none of the factors were significant predictors of gain.

Next, a null model was fit to quiz gain scores. An examination of the fixed effects indicated that the intercept was positive and statistically significant, see table 2 (Quiz Gain). The model revealed that the proportion correct on the quiz increased by 9 percent from pre- to posttest. Finally, an examination of the random effects revealed that a large proportion of the variability in gain scores was between classrooms. Specifically, the ICC was .18, which indicated that 18% of the variability in gain scores was between classrooms. A further analysis, regressing quiz gain scores on the control variables revealed that none of the factors were significant predictors of gain.
Table 2.
Summary of Fixed Effects for Self-Efficacy Gain Scores, Quiz Gain Scores, iSTEM Gain Scores and Participation in Poster Presentation from Final Models

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Self-Efficacy Gain</th>
<th>Quiz Gain</th>
<th>iSTEM Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)  t (df)</td>
<td>b (SE)  t (df)</td>
<td>b (SE)  t (df)</td>
</tr>
<tr>
<td>Model for Key Outcomes (β₀j)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (γ₀₀)</td>
<td>0.64 (4.89)</td>
<td>0.09 (3.58)</td>
<td>-0.01 (-0.44)</td>
</tr>
<tr>
<td></td>
<td>(.13)  (17)**</td>
<td>(.03)  (17)**</td>
<td>(.03)  (17)</td>
</tr>
<tr>
<td>iSTEM – Pretest (γ₁₀)</td>
<td></td>
<td>-0.26 (-5.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.05)  (280)***</td>
<td>(.05)  (280)***</td>
<td></td>
</tr>
<tr>
<td>iSTEM² - Pretest (γ₂₀)</td>
<td>0.14 (2.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.05)  (280)**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Level-1 predictors (iSTEM & iSTEM²) were group mean centered and Level-2 predictor (MEAN iSTEM) was grand mean centered.
* p < .05, ** p < .01, *** p < .001
A series of models were conducted to formally test hypothesis-two. First, a null model was fit to iSTEM gain scores. The ICC revealed that little between class variability in iSTEM scores (ICC = \( \frac{.00007}{.00007+.25} = .0003 \)), which indicated that less than 1% of the variability in gain scores was between classrooms. Second, the non-linear hypothesis was tested by regressing iSTEM gain scores on iSTEM and iSTEM\(^2\) pretest scores. As expected, iSTEM pretest was significantly and negatively associated with gain scores, while iSTEM\(^2\) was significantly and positively associated with gain scores. To make these finding more concrete, model predicted iSTEM gain scores were plotted against pretest iSTEM scores, see figure 3. As shown in the figure, low pretest iSTEM scores were associated with higher gain, while moderate to high pretest iSTEM scores exhibited no gain. A further analysis, regressing iSTEM gain scores on the control variables revealed that none of the factors were significant predictors of gain.

**Figure 3.** Model Predicted iSTEM Gain Scores for Students from Low to High iSTEM Perceptions at Pretest (with 99% Confidence Bands)

Note: Solid line indicates model predicted iSTEM gain score; dashed lines indicate the upper and lower limit of the 99% confidence band.

**Discussion**

Pre-collegiate students engaged in long-term team-based engineering design problems that emphasize, inquiry, design, testing, and making activities is a natural platform for the integration of iSTEM content into classrooms. In the current study, we report on an engineering design based interventions that bridges the gap between the new integrated science standards that include engineering design and well researched models of teaching and learning from the cognitive sciences. One of the primary goals of this project was to transform student understanding of the co-dependent nature of STEM content knowledge; however, the measurement of such knowledge has been elusive. Therefore, the current study reports on the
continued efforts of this research team to develop valid and reliable scales to measure integrated
STEM content knowledge.

A key finding from this study indicated that the measurement of student perceptions were
consistent with our expectations. Specifically, we had hypothesized that our instrument would
measure a single factor related to perceptions of the interrelated nature of STEM content
knowledge and our analysis supported that hypothesis. A single factor emerged from the EFA to
explain the pattern of student responses. A second psychometric validation was conducted on the
posttest data using confirmatory factor analysis. Once again, the results indicated that the single
factor model provided the good fit to the data. Together, these findings provide incremental
evidence of the structural validity of the scale, as well as evidence for the stability of the
structural model over time.

A second, more important, finding from the current study concerned the positive change in
student perceptions of the co-dependent nature of STEM content knowledge. We addressed this
issue by comparing student perceptions at pretest with their perceptions at posttest. Our analysis
indicated that overall, students did not exhibit significant change in their perceptions; however,
probing our data further indicated a surprising pattern of results. More specifically, our data
indicate that the intervention was most effective with students who started with low perceptions
of the interrelated nature of STEM knowledge. It appears consistent across two studies
carried out by the authors that long-term participation in an authentic engineering design problem
cultivated connections for those students who initially saw the fewest connections and benefits of
STEM content knowledge. Although far from conclusive, this finding is promising in that
interventions such as this may help to spur student understanding of the utility of, interest in, and
engagement in STEM knowledge for those students who are initially the least engaged. Future
studies of similar interventions should closely examine the initial STEM connections at different
levels of the continuum.

These results of replicating the psychometric validation of the instrument advances our
knowledge of desperately needed tool design within STEM education. Ultimately, such tool
design is required to accurately detect and measure levels of content connects and knowledge
fusion occurring when learning-in-doing through engineering design. Furthermore, the results
support our optimism for the future potential of being able to drill down to specific learning
outcomes across the STEM disciplines. The work required to realize this level of scale
specificity will require a blueprint of specific scientific and engineering practices, crosscutting
concepts, and core ideas (content standards and learning benchmarks).

The design of this study (i.e., cluster non-randomized single-group pretest-posttest design) limits
the inferences we can make about the efficacy of the intervention as well as the generalizability
of the findings. Further, the observed changes to initially low scoring students’ perceptions of the
connectedness of STEM disciplines could also be explained by regression to the mean.
Therefore, follow up studies are being conducted that further explore these relationships and
broaden the replicability and generalizability of the findings.

References

This program is based upon collaborative work supported by a National Science Foundation Grant No. 0841259; Colorado State University, Thomas W. Chen, Principal Investigator, Michael A. de Miranda and Stuart Tobet Co-Principal Investigators. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Appendix A. Sample content blueprint.

**Blueprint for: General Physics – Work and Energy Unit**

**Broad Concept:** Work and energy as tools to describe systems.

<table>
<thead>
<tr>
<th>Specific Concepts</th>
<th>Operationalization</th>
<th>Level of Blooms</th>
<th>Misconceptions</th>
<th>Item Number(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Define and explain energy and work</strong></td>
<td>1. Differentiate among gravitational potential energy, elastic potential energy and kinetic energy. 2. Explain what work is and why work communicates more information than force alone.</td>
<td>1. Comp. 2. Comp.</td>
<td>Considering each type of energy to be unrelated to each other. Considering work to be a vector because it is the product of 2 vectors.</td>
<td>1 – 3</td>
</tr>
<tr>
<td><strong>Conservation of energy</strong></td>
<td>1. Calculate gravitational potential energy, elastic potential energy and kinetic energy. 2. Mathematically demonstrate energy conversion from initial circumstances to final circumstances. 3. Calculate net work done on a system.</td>
<td>1. Appl. 2. Appl. 3. Synth.</td>
<td>Confusion between the two potential energy formulas. Omitting forms of energy from one or both sides of the equation. Omitting negative signs for work acting against motion (ie. friction).</td>
<td>6 7 - 8 9 - 11</td>
</tr>
<tr>
<td><strong>Relationships among energy, work, and power</strong></td>
<td>1. Calculate power. 2. Analyze force time graphs to determine work.</td>
<td>1. Appl. 2. Anal.</td>
<td>Misinterpreting the meaning and/or implications of power. Confusion regarding the units and scale on a force vs. time graph.</td>
<td>12, 15 13 – 14</td>
</tr>
</tbody>
</table>
Appendix B. Sample assessment items associated to specific content concepts.

1. An object of mass $m$ slides down an incline plane as shown in the image below. When the object reaches the bottom of the incline it compresses a spring completely. At the moment of full compression, which forms of energy are non-zero values?

I. Gravitational potential energy
II. Elastic potential energy
III. Kinetic energy

a. I only  
 b. II only  
 c. III only  
 d. II and III only  
 e. I, II, and III

2. A skier descends a mountain as shown below. When the skier is halfway down, which one of the following statements is true? Assume ideal conditions and negligible friction.

a. Potential energy $>$ kinetic energy  
 b. Potential energy $=$ kinetic energy  
 c. Potential energy $<$ kinetic energy  
 d. All of the above  
 e. None of the above

3. The work-energy theorem states that work is equal to…

a. a change in kinetic energy.  
 b. a change in potential energy.  
 c. a change in thermal energy.  
 d. the energy required to compress a spring.  
 e. any change in energy.
4. A skier descends a mountain as shown below. When the skier reaches the bottom of the mountain, how many times faster is she traveling as compared to when she was exactly halfway down?

\[ v_f = v_i^2 \]

b. \( v_f = 4 \times v_i \)

c. \( v_f = 2 \times v_i^2 \)

d. \( v_f = \sqrt{2} \times v_i \)

e. Cannot determine with given information

5. Which of the following are vector quantities?
   I. Work
   II. Velocity
   III. Displacement

a. I only
b. II only
c. III only
d. II and III only
e. I, II, and III

6. A spring (\( k = 650 \text{ N/m} \)) is compressed 0.22 m from its equilibrium position. How much energy is stored in the spring?
   a. 15.7 J
   b. 23.7 J
   c. 31.4 J
   d. 71.5 J
   e. 5110 J

7. Consider a 5.4-kg ball being dropped from a height of 1.8 m. After the ball has fallen the entire 1.8 m it lands on a spring (\( k = 750 \text{ N/m} \)). How much does the spring compress? (assume that the spring can accommodate the required compression)
   a. 0.143 m
   b. 0.289 m
   c. 0.330 m
   d. 0.504 m
   e. 0.579 m
8. A skier is traveling at 5.5 m/s at the top of a hill that is 18 m above a valley below. How fast is the skier traveling when she reaches the valley? (assume that friction is negligible)
   a. 4.04 m/s
   b. 11.1 m/s
   c. 19.6 m/s
   d. 21.4 m/s
   e. 49.5 m/s

9. How much work is required to stop a 1500-kg car that is traveling at 12.5 m/s?
   a. 117 KJ
   b. -117 KJ
   c. 189 KJ
   d. -189 KJ
   e. None of these.

10. A 550 N crate rests on the floor. How much work is required to move it 2.0 m along the floor against a friction force of 150 N at constant speed?
    a. 300 W
    b. 150 W
    c. 300 J
    d. 150 J
    e. 1100 W

11. A 550 N crate rests on the floor. How much work is required to move it 2.0 m vertically at constant speed of 3 m/s?
    a. 110 W
    b. 110 J
    c. 1100 W
    d. 1100 J
    e. None of the above

12. A 57 kg person climbs to the top of a 5 m ladder in 3 seconds. How much power did she generate?
    a. 950 W
    b. 950 J
    c. 2850 W
    d. 2850 J
    e. None of the above
13. How much work is done by the applied force in moving the object from 0 m to 0.75 m?

![Applied Force Graph]

- a. 0.80 J
- b. 0.45 J
- c. 0.60 J
- d. 0.64 J
- e. 0.25 J

14. How much work is done by the applied force in moving the object from 0 m to 0.50 m?

![Applied Force Graph]

- a. 0.80 J
- b. 0.45 J
- c. 0.60 J
- d. 0.64 J
- e. 0.25 J
15. Find the power of a man who pushed a box horizontally 8 m with a force of 15 N in 6 seconds. Assume that friction is negligible.
   a. 20 W
   b. 120 W
   c. 20 J
   d. 120 J
   e. There is not enough information to determine power.

Answers
1. B
2. B
3. A
4. D
5. D
6. A
7. D
8. C
9. B
10. C
11. D
12. A
13. B
14. E
15. A
Appendix C. Content validation instrument.

[Name:________________________ Teacher:___________________ Subject:____________________]

Thank you for agreeing to serve as an expert in this content validation process! You have been asked to participate in this process because of your specialized knowledge regarding content covered in the projects. For that reason, I ask that you help to judge the items on the teacher made quizzes. I would appreciate your help in evaluating whether the following items measure what they are designed to measure.

Please begin the validation process by reading each of the specific concepts in the attached blueprints. Then fill-in the Conceptual Definitions (below) with each of the Specific Constructs covered in the blueprint.

**Note:** the number of constructs will vary from project to project.

<table>
<thead>
<tr>
<th>Construct Categories</th>
<th>Conceptual Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td></td>
</tr>
<tr>
<td>V.</td>
<td></td>
</tr>
</tbody>
</table>

VI. None of the above

The item does not seem to fit into any of the categories.

**Rating Tasks**

1. Construct Rating: Please place each item into one of the above six construct categories.
2. Confidence Rating: Please rate HOW CERTAIN you are that the item fits into the construct.
3. Relevance Rating: Please rate how RELEVANT/APPROPRIATE the item is to the construct (i.e., how representative of the construct is the item).
4. Double check the correct answer.
5. Identify any distracters that do not align with “misconceptions” (listed in the blueprint) or are not relevant to the construct.
**Directions**

This quiz is being conducted to better understand students’ understanding of the concepts/content used in the project.

Place an “x” for in the box that best describes your evaluation of the Construct, Confidence, and Relevance ratings.

Finally, identify distracters (wrong answers) that do not take advantage of “misconceptions” or are otherwise not relevant.

<table>
<thead>
<tr>
<th>Construct Categories</th>
<th>Confidence Rating</th>
<th>Relevance Rating</th>
<th>Correct Answer</th>
<th>Distracter Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>None of the above</td>
<td>Low certainty</td>
<td>Low relevance</td>
<td>Mark as Yes or No</td>
</tr>
<tr>
<td>II.</td>
<td></td>
<td>Moderate certainty</td>
<td>Moderate relevance</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td></td>
<td>Complete certainty</td>
<td>Complete relevance</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

List distracters that are not aligned with “misconceptions” or are not relevant.

---

Item-1

Item-2

Item-3

Item-4

Item-5

Item-6

Item-7

Item-8

Item-9

Item-10

Item-11

Item-12

Item-13

Item-14

Item-15

Do you have any comments about the items or the constructs?

Is anything that you would add to this quiz?