AC 2011-435: MEASURING THE EFFECTS OF INTEGRATING ENGINEERING INTO THE ELEMENTARY SCHOOL CURRICULUM ON STUDENTS' SCIENCE AND ENGINEERING DESIGN CONTENT KNOWLEDGE

Melissa Dyehouse, Purdue University

Melissa Dyehouse is a Postdoctoral Research Associate at the Institute for P-12 Engineering Research and Learning (INSPIRE). She received her M.S.Ed. and Ph.D. in Educational Psychology from Purdue University. She has conducted research on instrument development and validation, students' perceptions of engineers and scientists, and the effectiveness of interventions to improve students' perceptions and attitudes about science, technology, engineering, and math (STEM) disciplines. Her research at INSPIRE focuses on the learning and teaching of engineering as a "caring" discipline in the context of environmental and ecological concerns.

Heidi A. Diefes-Dux, Purdue University, West Lafayette

Heidi Diefes-Dux is an Associate Professor in the School of Engineering Education at Purdue University. She received her B.S. and M.S. in Food Science from Cornell University and her Ph.D. in Food Process Engineering from the Department of Agricultural and Biological Engineering at Purdue University. Since 1999, she has been a faculty member in Purdue's First-Year Engineering Program, the gateway for all first-year students entering the College of Engineering. She is currently the Director of Teacher Professional Development for the Institute for P-12 Engineering Research and Learning (INSPIRE). Her P-12 research interests center on the integration of engineering into elementary education.

Brenda Capobianco, Purdue University

Dr. Brenda Capobianco is an Associate Professor in the Department of Curriculum and Instruction and holds a courtesy appointment in the School of Engineering Education and an affiliated appointment in Women's Studies at Purdue University. She holds a B.S. in biology from the University of Alaska Fairbanks, M.S in science education from Connecticut Central State University, and Ed.D. from the University of Massachusetts Amherst. Her research interests include girls' participation in science and engineering; teacher's engagement in action research; and science teachers' integration of the engineering design process to improve science learning.

Measuring the Effects of Integrating Engineering into the Elementary School Curriculum on Students' Science and Engineering Design Content Knowledge

Abstract

Engineering education and engineering education teacher professional development programs are becoming more common at the elementary level. However, there are no well-developed instruments designed to assess the impact of students' knowledge in subjects that engineering is thought to improve, such as the engineering design process, science, and technology. To address this need, student knowledge tests for grades two, three, and four were developed to measure knowledge of engineering design, science, and technology, both before and after an engineering curriculum was implemented by teachers who attended an engineering education professional development program. This paper reports the progress and methods of two iterations of instrument development for the student knowledge tests. A total of 386 students in grades 2 to 4 from the year 2008-09 cohort and 636 students in grades 2 to 4 from the year 2009-10 cohort participated in the study. Based on item analysis results, the tests were revised and administered to a new group of students (2009-10 cohort). Item analysis of the pre-items flagged several potential revisions dependent upon forthcoming results of the post item analysis for year two. Implications include redesign of the student knowledge test for the academic year 2010-11, and the inclusion of more items for engineering and technology subscales to improve reliability. Additional implications include support for the use of engineering as an integrative context for science and engineering learning in K-12 classrooms.

Introduction

It is impossible to overstate the importance of developing assessments that measure student knowledge gains upon completion of a teacher professional development program to incorporate engineering into the elementary classroom. Engineering curricula and engineering teacher professional development at the elementary level remains a developing area[1]. It follows that assessments measuring the impact of such teacher professional development programs, or engineering interventions on students' engineering design, science, and technology knowledge, have not been widely developed or utilized. For example, the National Academy Engineering (NAE)[1] reports that there is a "paucity of data" available to assess the impacts of K-12 engineering education on many student outcomes, which "reflects a modest, unsystematic effort to measure, or even define, learning and other outcomes" (p. 154).

There is a need for assessments that are developmentally appropriate and easily administered to students in grades 1 through 5. One challenge in developing such assessments is the need for an assessment for large numbers of students that may be affected by a teacher professional development program. Although multiple measures should be used to assess impact, a quantitative measure of student knowledge that can be administered quickly by all teachers in both treatment and control classrooms is necessary in order to have a standard comparison across a potentially large number of treatment and control group classrooms. Additionally, the assessments must be developmentally appropriate for students at the elementary level. This may entail separate sets of directions for administration (e.g., the teacher reads each question to students versus the students reading the questions themselves) or limiting the number of test items used depending on the age of the child.

Assessment of a program's impact is necessary for multiple stakeholders. These assessment results are important for administrators who want to know the program's impact on factors that will eventually assist state testing (e.g., science). Engineering education researchers want to determine whether students are learning about engineers and engineering as well as assess the impact of teacher professional development programs on students' engineering design, science, and technology knowledge. Teachers need to ascertain whether students understand the engineering lessons they implement as well as whether the lessons are a worthwhile use of their limited classroom time.

There is a need for a student knowledge test capable of accurately measuring students' knowledge of engineering design, science, and technology as a means to inform teacher professional development in order to aid engineering education at the elementary level.

Literature Review

Challenges of assessing the impact of engineering curricula at the elementary level

There are many challenges of assessing the impact of engineering curricula in grades 1 through 5. One of these challenges is elementary students' lack of familiarity with state testing or standardized test items. The majority of states begin standardized testing in 3rd grade.[2] This lack of familiarity at the elementary level, especially below 3rd grade, may have an effect on students' engineering-related assessment scores. For example, Callenbach[3] found that the standardized reading test scores of inexperienced 2nd grade test-takers improved significantly following instruction and practice with standardized tests.

Another issue at elementary level is the construction of items that are developmentally appropriate for students. For example, the upper limit of the number of items recommended for use on a test for students at the 1st grade level is relatively low, because of younger children's lower level of attention[4,5]. This low number of items poses a problem, however, because with fewer items, an instrument will become less reliable. Additionally, language must be adjusted to a level comprehensible to a student's respective grade level. At lower grades, for example, pictures of faces with various expressions (happy, neutral, sad) are often used for students to select the appropriate response to an item (e.g., McKenna & Kear[6]).

The NAE[1] recommends collecting data to examine the impact of student learning in STEM. For example, an NAE[1] study found that assessment of the impacts of engineering education on students' achievement and attitudes in K-12 was often an "afterthought;" rarely using valid pre-tests and often inquiring about enjoyment rather than measuring knowledge gains (p. 63). A challenge occurs due to a lack of rigorously developed instruments to measure engineering design, science, and technology learning after a new engineering curriculum is implemented. For example, the Engineering is Elementary (EiE) curriculum developed by the Boston Museum of Science contains lesson-specific questions to measure student knowledge gains, rather than more general questions on engineering design, science, and technology[5].

Purpose and Research Questions

The purpose of this study was to develop an instrument capable of measuring the impacts of a teacher professional development (TPD) program to integrate engineering into the elementary classroom on students' engineering design, science, and technology learning. Engineering is not commonly taught at the K-12 level, and as such there is minimal research on student learning outcomes. There is also a lack of instruments to measure the potential knowledge gains of a K-12 engineering curriculum and to better inform teacher professional development practices.

The research questions are:

Are the student knowledge tests reliable measures of students' learning in engineering design, science, and technology following an engineering curriculum?

Are the student knowledge tests capable of showing changes following the implementation of an engineering curriculum?

Participants

Students

Participants included 386 students in grades 2 through 4 from one school district in the south-central United States taking part in the study during the 2008-09 school year, and 636 students in grades 2 through 4 taking part in the study during the 2009-10 school year. Student participants represented ethnically diverse populations from both urban and suburban elementary schools, including ten participating classrooms from one school district in the south-central United States. Tables 1 and 2 show the breakdown of student participants by sex, Title 1 status, test/control group status, and ethnicity according to the students' grade both before and after the test in cohorts 2008-09 and 2009-10.

Table 1. Student demographics for 2008-09.

Test/Grade	**Sex**		**Title 1**		**Test/Control**		**Race**			
	M	*F*	*Title 1*	*Not Title 1*	*Treated*	*Control*	*Asian/ Pacific Islander*	*African Amer.*	*Hisp.*	*White*
Grade 2 pre	28	23	20	31	28	23	6	6	14	22
Grade 2 post	30	25	25	30	27	28	9	8	16	22
Grade 3 pre	41	39	47	33	49	31	4	21	22	20
Grade 3 post	33	33	37	29	37	29	4	19	21	14
Grade 4 pre	39	38	46	31	42	35	6	15	27	22
Grade 4 post	29	28	37	20	33	24	5	12	22	17
Total	200	186	212	174	216	170	34	81	122	117

Table 2. Student demographics for 2009-10.

Test/Grade	Sex		Title 1		Test/Control		Race			
	M	*F*	*Title 1*	*Not Title 1*	*Treated*	*Control*	*Asian/ Pacific Islander*	*African Amer.*	*Hisp.*	*White*
Grade 2 pre	81	88	56	103	159	10	26	36	52	51
Grade 3 pre	119	110	97	82	179	50	19	51	88	66
Grade 4 pre	108	122	84	123	207	24	22	33	75	97
Total	308	320	237	308	545	84	67	120	215	214

Teachers

The number of teachers associated with treatment and control groups are 11 teachers in the treatment group and nine teachers in the control group for the 2008-09 cohort and 38 teachers in the treatment group and 6 teachers in the control group for the 2009-10 cohort. The number of years that both teacher cohorts had been in the classroom ranged from 1 to 35 with a mean of 10. Teachers had no prior learning experiences with engineering.

Instrument

For the 2008-09 year of implementation, the student knowledge tests contained a total of eleven questions (2nd grade), ten questions (3rd grade), and sixteen questions (4th grade) organized into three domains of knowledge: science related content, engineering (e.g., the engineering design process, the work of an engineer). The tests were composed of developmentally appropriate multiple-choice and open-ended items that probe for different levels of comprehension using low, medium, and high cognitive demand items. Items were generated by members of the research team including STEM faculty, research assistants, and elementary educators. Science content items from national and state-wide educational performance assessments were used. Items from state-wide assessments were used to ensure consistency in language and cognitive development. Engineering items were modeled after EiE unit assessment items.

Teacher Professional Development with Engineering

In June 2008, a week-long summer academy was conducted at one of the elementary schools on the district site for 32 elementary (grade 2-4) teachers representing seven elementary schools in the district. The main goals were to enable teachers to (1) convey a broad perspective of the nature and practice of engineering; (2) articulate the differences and similarities between engineering and science thinking; (3) develop a level of comfort in discussing what engineers do and how engineers solve problems with elementary students; and (4) use problem-solving processes (i.e., science inquiry, model development, and design processes) to engage their students in complex open-ended problem solving. In essence, teachers began to learn who engineers are, what engineers do, how to present complex problems to their students, and guide solution development through an engineering design process.

Some of the professional development materials drew on those of the Museum of Science in Boston. These materials focused on (1) developing a broad definition of technology, (2) introducing the engineering design process, and (3) learning about particular kinds of engineers, making connections to science content, and employing the engineering design process through select *Engineering is Elementary* (EiE) units. Additionally, teachers were exposed to elementary level mathematical modeling problems. Teachers participated in all activities as students, and then reflected on and discussed how these activities would be implemented in their classrooms. Special events offered during the academy included an opening reception to get to know one another and an industrial engineering focused tour of a local pet carrier manufacturing company. The local university hosted and co-sponsored a K-12 and Engineering Education Dinner and provided practicing engineer panel speakers. The university also provided a tour of their competitive automotive design facilities for student teams.

In August 2009, 23 of the original 32 teachers returned for a second three-day academy. This academy was designed to address teachers' concerns and continued professional development interests expressed at the end of 2008-09 in interviews and issues noted by the research team through student and teacher data collection and analysis of various types throughout the year. Teachers desired to learn how to better handle teaming in their classrooms, debrief with other teachers at the same grade level on the implementation of their lessons, and observe some new engineering activities. The research team wished to address minimal use of engineering notebooks, lack of discussion about the practice of engineering around activities targeted to specific disciplines, missed opportunities to incorporate mathematics and science principles, labeling activities as "engineering activities" while missing key features, and lack of assessment of student learning from the engineering activities. The objectives of this academy were to enable teachers to: (1) identify opportunities to augment science or mathematics learning through engineering, (2) comfortably discuss what engineers do and describe select types of engineering, (3) use engineering design process and model development process to engage elementary students in complex open-ended problem solving, and (4) assess student learning across multiple dimensions through engineering activities.

Teachers spent considerable time in this academy debriefing the 2008-09 implementation of the EiE units, walking though the EiE unit, and discussing the many lessons learned. Two new activities were implemented to enable teachers to recognize when an activity is an engineering design activity and when it is not. The opportunities for exploring math and science concepts through these activities were also highlighted. Assessment strategies were discussed in association with each of these activities. To bolster teachers' understanding of the field of engineering depicted in their grade level's respective EiE unit, teachers used internet resources to assemble content for a poster or handout about their particular field of engineering focusing on the actions of that particular kind of engineer, the resources that engineer uses, and the people that interact with that engineer. A session on teamwork with emphasis on building a classroom code of cooperation and a classroom agreed-upon set of rules of behavior for teaming activities was also included in this academy.

Also in August 2009, a new group of 36 elementary (grade 2-4) teachers from the district attended the week-long version of the academy. These teachers represented nine elementary schools that applied, including five schools new to this project. This group of teachers had the opportunity to interact with the original 2008-09 group of teachers. The 2008-09 group presented a list of lessons learned and talked about their experiences, and the new group got to ask questions.

Grades 2-4 Engineering and Science Content

At the district level, a commitment was made for each grade level to teach one EiE unit. The EiE unit was selected based on the relevance of its science content to the science content already being covered at the grade level. In short, through each unit, students learn about a type of engineering, connect science concepts to an engineering design project through an investigation activity, and employ the EiE engineering design process to complete the engineering design project. Grade two adopted “A Work in Process: Designing a Play Dough Process” (Solids & Liquids; Chemical Engineering)[7], grade three adopted “Marvelous Machines: Making Work Easier” (Simple Machines; Industrial Engineering)[8], and grade four adopted “Thinking Inside the Box: Designing a Plant Package” (Basic Plant Biology; Packaging Engineering)[9]. In addition to these lessons, many teachers taught a lesson centered on defining technology and a lesson that introduced the engineering design process by either verbally walking through the process to learn the steps or completing a short design project (e.g., index card tower). These were typically taught prior to the EiE unit. Some teachers opted to teach additional short design projects before and after the EiE unit.

Research Design and Procedures

A quasi-experimental design with pre-post tests for treatment and control groups was used to study the impact of the teacher professional development engineering education program on students’ knowledge by using the student knowledge test as the outcome measure.

The treatment participants were the 2nd, 3rd, and 4th grade students whose teachers had received engineering teacher professional development as described above and who, upon returning to their classrooms in the fall, agreed to teach the engineering lessons/curriculum they had learned in the academy. The control participants were 2nd, 3rd, and 4th grade students whose teachers had no engineering teacher professional development and would not teach engineering lessons in their classrooms that academic year. Treatment teachers agreed to administer the knowledge test both before any engineering instruction took place (pre) and after all engineering instruction had taken place (post). Control teachers administered the instrument in early Fall (pre) and late Spring (post).

Data Analysis

To determine whether the student knowledge test was able to show significant changes from pre to post, an ANCOVA was conducted for each grade level using the post engineering questions as the dependent variable and the pre engineering items as the covariate while also examining the effect of several other variables (independent variables include treatment/control group, sex, Title 1 status, and ethnicity) on students’ knowledge scores.

Item analysis methodology

Item analysis was conducted to determine internal consistency reliability of the post knowledge test items for each subscale. Item analysis was carried out using the treatment group of students.

Item difficulty

Item difficulty and discrimination values were calculated for the all of the items. Difficulty values (p) are the proportion of examinees that answer an item correctly[10]. These values may range from 0.00 to 1.00, with higher p values indicating an easier item. Items that are too easy on the pre-test may require revision because they will not be able to show any post changes. This is the main consideration for this analysis of the pre test difficulty levels. There should not be many difficult items on the post test. If less than half of students correctly answer an item on the post test, this indicates there may be issues either with the instruction (e.g., not covered, not covered well enough) or the item itself (e.g., confusing, too difficult to understand, not applicable to what students learned, poorly written).

Item discrimination

Item discrimination indices are used to estimate the extent to which success on an item corresponds to success on the whole measure. More specifically, discrimination is the extent to which items are discriminating between students with high knowledge and low knowledge as measured by the total knowledge score[10]. Item discrimination was calculated for each item using a point biserial correlation (ρ_{pbis}), which is a simplified computational formula for the Pearson product moment coefficient. The point biserial correlation represents the correlation between an item score and the total score with that item removed. A positive correlation is indicative of a correctly functioning item. The higher the value, the more proficient the item discriminates. A negative or zero value indicates that the item shows little or no discrimination and should be considered for revision.

Internal consistency reliability

Internal consistency is a form of reliability that indicates if items are measuring the same underlying construct[10]. Cronbach's α is a commonly utilized statistic that calculates the inter-correlations between these items and serves as a measure of internal consistency reliability. Cronbach's α ranges from 0.00 to 1.00. The commonly accepted cutoff for items to form a reliable scale is ≥ 0.70.

Item Analysis Results

Internal consistency reliability was adequate for all of the test items for each of the 2nd, 3rd, and 4th grade knowledge tests, respectively ($\alpha = 0.87$, $\alpha = 0.69$, $\alpha = 0.73$). However, the individual subscales of engineering and science showed inadequate reliability, possibly due to a lack of items.

For the 2nd grade knowledge test, there was only one engineering item, so reliability was not calculated for an "engineering" subscale. The reliability for the "science" subscale was calculated at $\alpha = 0.90$, which is considered high. Finally, when all of the items were combined, the overall reliability was $\alpha = 0.87$, which is good internal consistency reliability.

For the 3rd grade knowledge test, the reliability of the "engineering items" was low ($\alpha = 0.30$). The reliability for the "science" subscale was moderate, at $\alpha = 0.64$. The internal consistency reliability for the whole test was $\alpha = 0.69$, which is nearly acceptable to function as a scale.

Finally, for the 4th grade knowledge test the reliability of the "engineering" subscale was low ($\alpha = 0.35$). The "science" items showed moderate reliability ($\alpha = 0.63$) while the overall reliability for all items was $\alpha = 0.73$, which is acceptable to function as a scale.

Group Comparison Results

One method of examining the effectiveness of an assessment instrument is to determine if it captures pre-post differences[11]. To address the question of whether the instrument was capable of capturing pre-post differences, we conducted an analysis of covariance (ANCOVA) for each grade level using the post engineering questions as the dependent variable and the pre engineering items as the covariate while also examining the effect of several other variables (independent variables include treatment/control group, sex, Title 1 status, and ethnicity) on students' knowledge scores.

Engineering

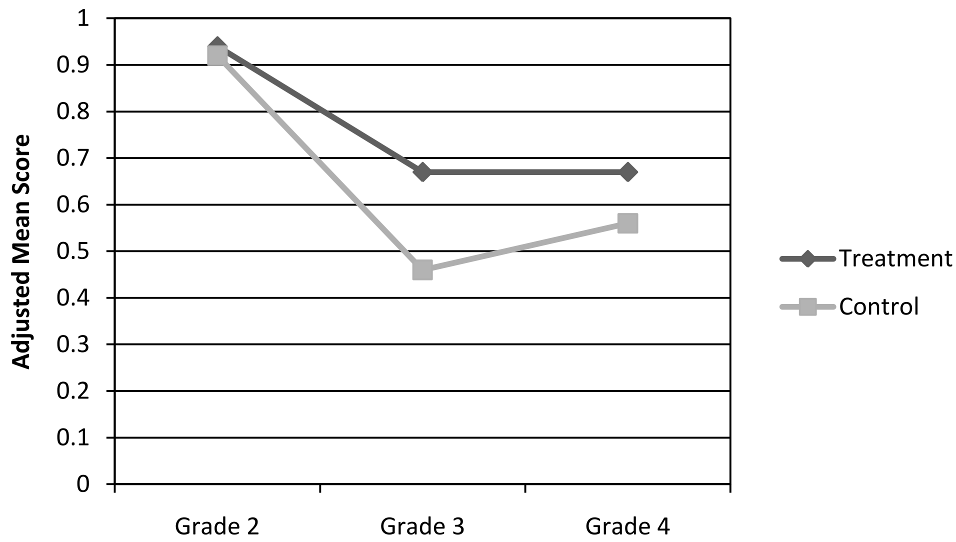
The 2nd grade engineering items (engineering) were analyzed separately for significance. Results of the ANCOVA showed no significant differences between pre and post scores, nor did any variables have a significant effect on students' knowledge scores. Concerning the 3rd grade knowledge test (engineering), there were no significant differences between pre and post scores, nor did any variables have a significant effect on students' knowledge scores. For the 4th grade knowledge test (engineering), there were significant effects of Title 1 status on students' engineering knowledge. This is to say that students who were *not* in a Title 1 school achieved increased scores (adjusted mean = 0.81) on the engineering items as compared with students in a Title 1 school (adjusted mean = 0.60). However, no treatment or control group differences were found.

Science

For the 2nd grade knowledge test (science), there were no significant differences between pre and post scores, nor did any variables have a significant effect on students' knowledge scores. Results of the 3rd grade knowledge test (science), showed no significant differences between pre and post scores nor did any variables have a significant effect on students' knowledge scores. Finally, for the 4th grade knowledge test (science), there were significant effects on the treatment group. Results showed that students who were in the treatment group achieved increased scores (adjusted mean = 0.66) on the science items as compared with students in the control group (adjusted mean = 0.51).

Total knowledge score

While there were no significant differences between 2nd grade treatment and control groups, there were statistically significant effects ($p < .05$) on treatment groups of 3rd and 4th grade students' total knowledge scores (Figure 1). More specifically, for 3rd grade knowledge test scores, students who were in the treatment group achieved increased scores (adjusted mean = 0.67) as compared with students in the control group (adjusted mean = 0.46). For 4th grade knowledge test scores, students who were in the treatment group achieved increased scores (adjusted mean = 0.67) as compared with students in the control group (adjusted mean = 0.56).



Note. Scale is from 0 to 1.
Figure 1. Post adjusted means on knowledge scores for treatment and control groups.

For the 2nd grade knowledge test scores, there were no significant differences between pre and post scores, nor did any variables have a significant effect on students' knowledge scores.

For the 3rd grade knowledge test there were significant effects on treatment group of students' overall knowledge score. Results showed that students who were in the treatment group achieved increased scores (adjusted mean = 0.67) as compared with students in the control group (adjusted mean = 0.46).

For the 4th grade overall knowledge test score there was one significant interaction between treatment group and Title 1 status, which indicates that students in a treatment or control group performed significantly different on the post knowledge test depending on whether they were in a Title 1 school or a non Title 1 school. Students who were in the control group scored lower if they were not in a Title 1 school than students who were in a control group but in a Title 1 school. Students who were in the treatment group scored lower if they were in a Title 1 school than if they were not in a Title 1 school. Results also showed significant effects of treatment group on students' overall knowledge scores. Specifically, students who were in the treatment group achieved increased scores (adjusted mean = 0.67) as compared with students in the control group (adjusted mean = 0.56).

Year 2 Revisions

Based upon the item analysis for the 2008-09 year, revisions were made to the student knowledge tests for the 2009-10 year without performing additional levels of item analysis (i.e., discrimination and difficulty). The biggest issue was low reliability for the engineering subscale of the test for each grade level. To correct this, more engineering-related items were added. Another issue was the lack of pre-post differences that were found for the 2nd grade knowledge test. Therefore, ambiguous item responses were eliminated and all items were revised so that all were multiple-choice in format. Finally, technology items were added to each of the tests so that this domain could also be measured.

The numbers of items in each subscale for grade two were seven science items, six engineering items, and two technology items. For grade three, there were seven science items, seven engineering items, and one technology item. Finally, the numbers of items in each subscale for the 4th grade test were seven science items, six engineering items, and two technology items. The pre-test results of the 2009-10 cohort are presented here.

Item Analysis Results for the Pre Student Knowledge Tests

Internal consistency reliability

Cronbach's α may be low if the test is short, items are measuring very different things, there is a small ability range in the sample, or the test contains only very easy or very difficult items. Because several of these factors may be present for these pre-knowledge items, the internal consistency reliability will be conducted with the post items to see whether α improves.

Table 3 shows the α values for 2nd, 3rd, and 4th grade subject areas. The "technology" subscale α value was not calculated for 3rd grade because there was only one item for this subject area.

Table 3. Cronbach's α for the knowledge test and subscales across grade levels.

Grade	All items	Science	Engineering	Technology
2nd	0.331	0.324	- 0.018	- 0.096
3rd	0.344	0.224	0.103	--
4th	0.643	0.442	0.467	0.252

All of these values are considered too low to be a reliable scale that measures students' knowledge of engineering design, science, and technology. However, if the items are able to reveal students' changes after the treatment occurs, the α values should increase.

Item difficulty

The difficulty of the 2nd grade knowledge items ranged from 0.05 to 0.80 (Table 4). Most of the items were below 0.50 in difficulty, suggesting these items are good for a pre-test where participants are assumed to have little knowledge about the topics. The five questions can also be considered acceptable because all are sufficiently difficult (> 0.80) and it is appropriate to begin the test with easier questions to minimize participant frustration. According to this analysis, the items may be ordered on the pre-test from least to most difficult, as displayed in Table 4 with the difficulty values for each item.

Table 4. Difficulty values for 2nd, 3rd and 4th grade knowledge test items.

Grade 2		Grade 3		Grade 4	
Item Number	**Difficulty Value**	**Item Number**	**Difficulty Value**	**Item Number**	**Difficulty Value**
Item 4	0.80	Item 1	0.63	Item 2	0.84
Item 1	0.78	Item 7	0.51	Item 3	0.72
Item 13	0.75	Item 12	0.40	Item 5	0.64
Item 11	0.65	Item 3	0.39	Item 9	0.63
Item 7	0.56	Item 13	0.39	Item 14	0.61
Item 12	0.36	Item 10	0.35	Item 10	0.54
Item 2	0.32	Item 2	0.34	Item 8	0.43
Item 3	0.29	Item 11	0.32	Item 13	0.39
Item 5	0.23	Item 14	0.23	Item 4	0.38
Item 9	0.19	Item 8	0.23	Item 12	0.26
Item 10	0.13	Item 9	0.23	Item 1	0.25
Item 8	0.12	Item 6	0.20	Item 6	0.25
Item 14	0.10	Item 5	0.19	Item 7	0.25
Item 15	0.10	Item 15	0.17	Item 15	0.23
Item 6	0.05	Item 4	0.13	Item 11	0.20

The difficulty for the 3rd grade knowledge items ranged from 0.13 to 0.63 (Table 4). Most of the items were below 0.50 in difficulty, which suggests these items are good for a pre-test where participants are assumed to have little knowledge about the topics. The easiest item is also acceptable because it is of moderate difficulty (0.63).

The difficulty for the 4th grade knowledge items ranged from 0.84 to 0.20 (Table 4). Most of the items were below 0.50 in difficulty, which again suggests that these items are acceptable for a pre-test of student knowledge. Table 4 displays the items from least to most difficult, with the difficulty values for each item (grades 2 – 4).

Item discrimination

Pre-survey discrimination values for the 2nd grade knowledge items ranged from – 0.001 to 0.32 (Table 5). The high level of difficulty for the items may have caused the low item discrimination of these items. If very low or negative discrimination values persist on the post-test knowledge items, the items should be revised or discarded.

Table 5. Discrimination values for 2nd, 3rd, and 4th grade knowledge items.

Item Number	ρ_{pbis}	**Item Number**	ρ_{pbis}	**Item Number**	ρ_{pbis}
Item 1	0.14	Item 1	0.16	Item 1	0.05
Item 2	0.13	Item 2	0.07	Item 2	0.24
Item 3	0.32	Item 3	0.17	Item 3	0.31
Item 4	0.20	Item 4	0.04	Item 4	0.15
Item 5	- 0.001	Item 5	0.15	Item 5	0.33
Item 6	- 0.05	Item 6	0.08	Item 6	0.10
Item 7	0.03	Item 7	0.15	Item 7	0.37
Item 8	0.03	Item 8	0.15	Item 8	0.36
Item 9	0.00	Item 9	0.10	Item 9	0.15
Item 10	0.20	Item 10	0.12	Item 10	0.29
Item 11	0.07	Item 11	0.28	Item 11	0.23
Item 12	0.03	Item 12	0.05	Item 12	0.45
Item 13	0.15	Item 13	- 0.01	Item 13	0.35
Item 14	0.02	Item 14	- 0.07	Item 14	0.21
Item 15	0.25	Item 15	0.17	Item 15	0.38

Pre-survey discrimination values for the 3rd grade knowledge items ranged from – 0.07 to 0.28 (Table 5). Again, the high level of difficulty for the items may have caused the low item discrimination. Pre-survey discrimination values for the 4th grade knowledge items ranged from 0.05 to 0.45, which shows a low level of discrimination (Table 5).

Discussion and Implications

The primary purpose of this study was to develop an instrument capable of measuring the impacts of a teacher professional development program to integrate engineering into the elementary classroom on students' engineering design, science, and technology learning. Results of the item analysis and group comparisons showed that improvement was needed for the first iteration of the student knowledge tests. Following the results of item analysis of the second iteration, the instrument showed potential on the pre-test items and several items were flagged for revision.

More specifically, the item analysis results indicated the need for the inclusion of more items for the engineering and technology subscales to improve reliability. For the pre student knowledge tests, item order can be changed so that more difficult items are placed at the end of the test to minimize student frustration and to increase the likelihood that the students will complete the test, which is especially important for younger students with shorter attention spans[4]. Finally, there is a need for more items measuring engineering content especially at the 2nd grade level so that items can form a more reliable scale. Removing some of the science items can aid in keeping the test at 15 items.

In summary, these scales have the potential to measure students' knowledge because the difficulty values for the pre-test were low enough that it will be possible to see changes.

Difficulty, discrimination, and α values should increase on the post-test. However, until reliability increases, the knowledge test can be analyzed using all 15 items rather than separately as subscales.

One implication of this study is the redesign of the student knowledge test for academic year 2010-11, with the aim of showing improvements in reliability and pre-post detection of differences among students' knowledge test scores. The post tests for the year 2009- will provide more results about how the tests are functioning.

Additional implications include support for the use of engineering as an integrative context for science and engineering learning in K-12 classrooms. As seen by the pre-post changes on 3rd and 4th grade students' total knowledge scores, the teacher professional development program has the potential to successfully have an impact on students' knowledge.

Finally, an important implication is the eventual impact upon teaching practices and teacher professional development following the results of the assessment. For example, if students do not improve in the "science" subscale from pre to post, then the professional development program can be modified to show teachers how to make more connections between engineering and science.

In conclusion, the student knowledge tests fill a needed gap in the assessment of teacher professional development programs in engineering education at the elementary school level. More engineering education curricula and professional development programs are taking place in grades 1 through 5[4] and such assessment is needed. Further revision and data collection will take place to ensure that the instrument is reliable and capable of showing pre-post changes following an engineering education curriculum implemented following a teacher professional development program.

Acknowledgement

This work was made possible by a grant from the National Science Foundation (DRL 0822261). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Bibliography

1. National Academy of Engineering. (2009). *Engineering in K-12 education: Understanding the status and improving the prospects*. National Academies Press: Washington, DC.
2. Toye, C., Blank, R. K., Sanders, N., & Williams, A. (2006). *Key State Education Policies on PK-12 Education: 2006*. Washington, D.C.: Council of Chief State School Officers.
3. Callenbach, C. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second-grade students. *Journal of Educational Measurement, 10(1)*, 25-30.
4. Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*, 48–71.
5. Lachapelle, C., & Cunningham, C. (2010). Assessing elementary students' understanding of engineering and technology concepts. American Society for Engineering Education (ASEE) annual meeting, Louisville, KY.
6. McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher, 43(9)*, 626-639.

7. Museum of Science, Boston. (2011). *A work in process: Designing a play dough process*. Retrieved from Museum of Science, Boston, Engineering is Elementary web site: http://www.mos.org/eie/20_unit.php
8. Museum of Science, Boston. (2011). *Marvelous machines: Making work easier*. Retrieved from Museum of Science, Boston, Engineering is Elementary web site: http://www.mos.org/eie/20_unit.php
9. Museum of Science, Boston. (2011). *Thinking inside the box: Designing a plant package*. Retrieved from Museum of Science, Boston, Engineering is Elementary web site: http://www.mos.org/eie/20_unit.php
10. Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Wadsworth Group/Thomson Learning: Belmont, CA.
11. Hsu, M., Cardella, M., & Purzer, S. (2010). *Assessing elementary teachers' design knowledge before and after introduction of a design process model*. American Society for Engineering Education (ASEE) annual meeting, Louisville, KY.