121st ASEE Annual Conference & Exposition
Indianapolis, IN
June 15-18, 2014
Paper ID #10494

Analysis of Design Process Knowledge Task Responses: Statistical Approaches to Uncover Patterns (Research)

Dr. Monica E Cardella, Purdue University, West Lafayette

Monica E. Cardella is an Associate Professor of Engineering Education at Purdue University and the Director of Informal Learning Environments Research for INSPIRE (the Institute for P-12 Engineering Research and Learning). She has a BSc in Mathematics from the University of Puget Sound and an MS and PhD in Industrial Engineering from the University of Washington. Her research focuses on: parents' roles in engineering education; engineering learning in informal environments; engineering design education; and mathematical thinking in engineering.

Ms. Ming-Chien Hsu, Purdue University, West Lafayette

Ming-Chien Hsu is a doctoral candidate of Engineering Education at Purdue University. She was an electrical engineer working on semiconductor devices before stepping into the realm of education research. Her current research explores and characterizes learning experiences in engineering such as design and interdisciplinary learning.

Dr. George D Ricco, Purdue University, West Lafayette

George Dante Ricco is a recent graduate of the School of Engineering Education at Purdue University and is on the job market. His work focuses on applying various research methods to address long standing anecdotal questions, ranging from ethnographic studies to hierarchical linear models. He was born in Kent, Ohio. He attended Walsh Jesuit High School, and instead of becoming a Jesuit, he decided to go to Case Western Reserve University in Cleveland to obtain his BSE in engineering physics (2002). He then spent a number of years on the beach at the University of California at Santa Cruz, receiving master's degrees in physics (2007) and earth and planetary sciences (2008) until emancipated by Prof. Matthew Ohland at Purdue University. He enjoys cycling, weightlifting, running, photography, volunteering at a number of organizations, and the untold intellectual pleasures provided by the study of Lagomorph physiology. He resides in Lafayette, Indiana, and in-between job interviews spends time with his Leporidae life partner, Rochelle Huffington Nibblesworth.

Analysis of Design Process Knowledge Task Responses: Statistical Approaches to Uncover Patterns

Introduction

While engineering design has been included as a key criterion in assessing undergraduate engineering programs for decades [1], it has more recently been recognized as a national science standard for K-12 curricula with the release of the Next Generation Science Standards in 2013 [2], which were created through the collaboration of 26 different states. At the state level, more and more states have been incorporating engineering concepts and engineering design into their standards. With this, there will be increasing need for: teacher professional development, curricula, and assessments related to engineering design for K-12th education.

There has been effort in laying out the pathways of design learning in the engineering education community. For example, Atman and her colleagues have compared the processes used by college freshmen, seniors, and professional engineers in order to investigate how design strategies change with design experience [3-5]. Crismond and Adams [6] summarized results from various design studies and compared design strategies used by beginning designers versus those by informed designers. Furthermore, Adams, Turns & Atman [7] suggested that designers might take some form of "trajectory" along candidate dimensions of design. Also, they provided different perspectives into investigating learning progressions, such as the design process perspective. Although these research studies focusing on how college engineering students and experts design can provide invaluable information and perspectives on what design education in K-12 might look like, they are not sufficient for understanding how K-12 students learn design. In our study, we explore ways to describe elementary students' conceptions and understanding of the engineering design process and what the pathways of design learning may look like for elementary students.

Theoretical Framework

In this work, we consider the idea of design learning progressions. The concept of a "learning progression" is described as an "empirically-grounded hypothesis of successively more sophisticated ways of thinking about a fundamental disciplinary idea and practice" [8, 9] as a way to guide instructions and learning goals. These hypotheses describe pathways students are likely to follow to master core concepts [9].With respect to the engineering design process, understanding students' learning progressions can enhance our understanding of how we might approach teaching the design process to 2nd graders, and how this might differ from what and how we teach 3rd graders, and 4th graders, etc. The framework of learning progressions arose as a call for more integrated science curriculum across grades that focus on important ideas, progressing complexity, and not discrete facts. The progressions not only focus on content but also on inquiry practices [10]. Also, carefully designed instruction and curriculum should be part of learning progressions, as the focus on learning progressions stems from seeing the need to integrate standards, curriculum, and assessment [8].

The existing empirical studies on learning progression present different ways of developing learning progressions. In the first approach, progressions are developed based on synthesis and analysis of existing research on the domain [11]. In the second approach, studies instead use cross-sectional studies across multiple grades. Such studies does not entail designed instructions but focus on the progression under current status-quo of teaching [12]. The third approach is based on what students across multiple grade levels are able to achieve after being given carefully-designed instruction on the topic [13, 14]. Existing research studies on learning progressions also present differences in the grain size, relationships to curriculum and instruction, and relationship to prior knowledge.

Our study is characterized by the following features in relation to the development of learning progressions. First, we understand that instruction and curriculum are an integral part of considering learning progressions. However, since we know very little about elementary students' knowledge and reasoning about design, the first step is for us to find out the status quo of understanding in order to make suggestions on targeted instruction and progressions. Second, we are employing a cross-sectional study to document students' development of knowledge and reasoning on design across multiple grades. Third, learning progressions focus on fundamental and generative ideas in a discipline, and design has been identified as an important concept in engineering learning [15]. Some might argue that designing involves procedural skills. However, reflective practice that involves conceptual understanding of the design space and problems necessitate the practice of design.

Research Questions

The focus of this paper is to look at differences across 2nd ,3^{rd,},and 4th graders' understanding of design. We ask the following questions:

> *Are there discernible differences in elementary students of different grade levels' understanding of the engineering design process? If so, what are these differences?*

If we are able to identify specific differences between different grade levels' understanding of design, this can help us to imagine an engineering design learning progression where we might focus on one aspect of the engineering design process in 2nd grade, a different aspect of the design process in 3rd grade, and another in 4th grade.

Data Collection

We adapted an assessment instrument focused on assessing college students' and professional engineers' understanding of the engineering design process [16] for the elementary school context. Additional discussion of the process we used to adapt and validate the task is presented elsewhere [17]. We conducted one-on-one interviews with second, third, and fourth graders at nine elementary schools in one suburban school district during the school years that began in the falls of 2009, 2010, and 2011. The breakdown of the number of students interviewed by grade and school year is listed in Table 1. Some of the students were interviewed both at the beginning and at the end of the year, and some of them proceeded to be interviewed in the next school year or two. In total, 919 data points were collected.

Table 1. Study participants by grade and year

		School Year			Total
		2009-10	2010-11	2011-12	
Grade	2	74	71	152	297
	3	75	65	160	300
	4	62	96	164	322
Total		211	232	476	919

During the interviews, the interviewer first described a fabricated design process with the aid of illustrations of a student, Chris, doing different tasks during the process of designing a container for an egg-drop contest (See Figure 1). The students were then asked open-ended questions of (i) what they thought was good about the process and (ii) what should have been done differently.

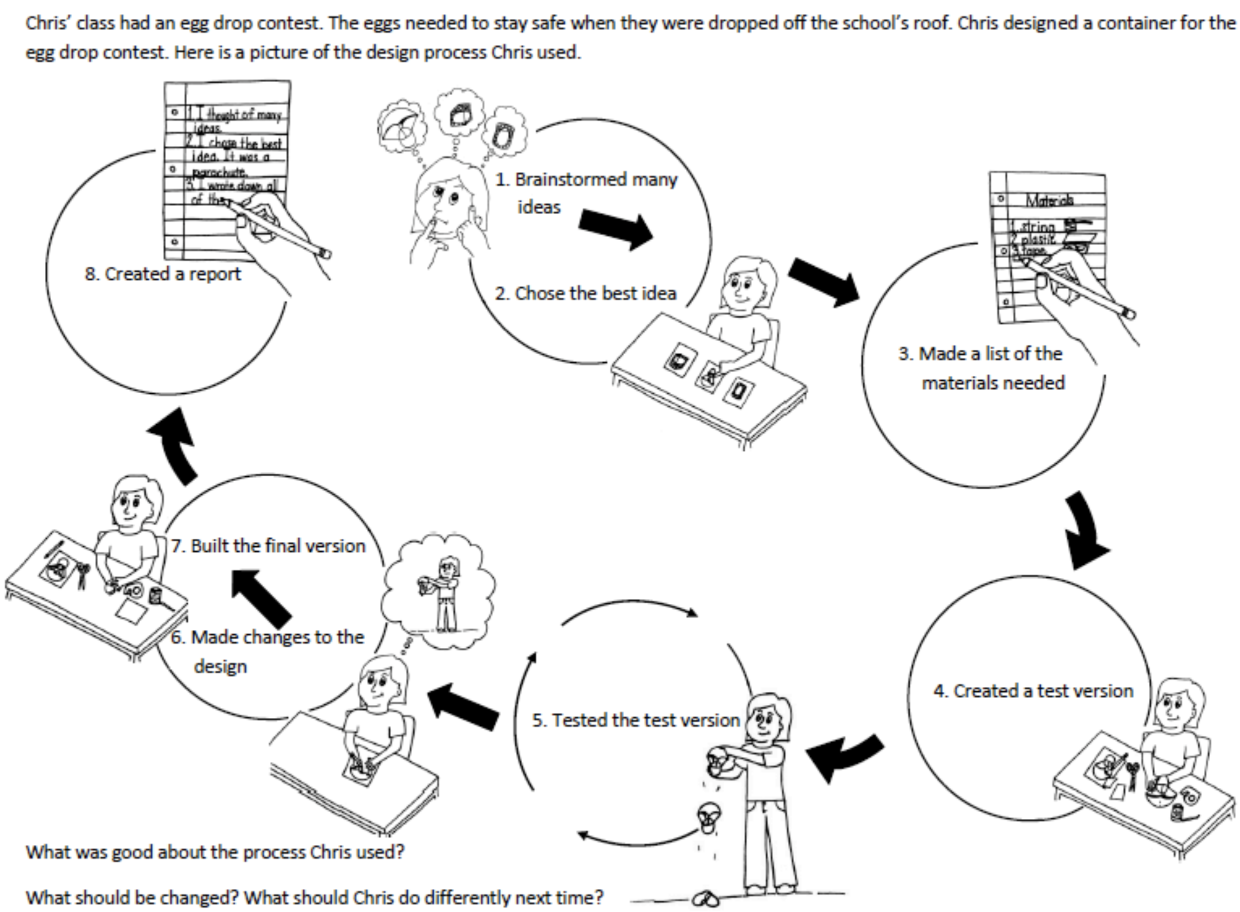

Figure 1. The instrument used to elicit students' responses

Data Analysis: Qualitative analysis and coding

Students' responses to the two questions were coded with eight design process knowledge concept codes developed in our earlier data collection [17]. The coding categories emerged from a process of grounded-coding coupled with comparing the codes that emerged to a version of the design process used in a popular elementary engineering curriculum, Engineering is Elementary [18]. Descriptions of the eight coding categories and example responses are included in Table 2.

Table 2: Definition and examples of coding categories

Concept/Coding Category		Category Explanation: Indicating that design process should include….	Examples of specific terms that students used
1	Ask	Asking about the details of the problem and constraints	• We asked questions about how is it going to make it more soft or is it going to be like a real egg
2	Imagine	Brainstorming ideas and picking a good idea	• He thought about it. Because if you think about it and drew it, it helps you better to pick which one and helps you do good. • He wrote down his He's brainstorming and trying very hard
3	Plan	Planning ahead, including the materials needed for finishing the design	• He said he what was going to before he started doing all this • He made a list of the materials he may need like a bucket
4	Create	Creating and building	• He created something • He built it differently
5	Improve	Making the design even better	• If it didn't work too well, she might want to make a few more changes than she did • He improved it • He was fixing his project he was redoing it to make it not break the egg
6	Test	Testing out the prototypes built	• You don't know if it works if you don't test them. • He tested the test version…. So he can see what he needs to add
7	Document	Taking notes of what ideas came up and what was done	• He wrote a report about it… So that ummm everybody else knows. • He's supposed to write what he think. Then if he forget, he can read his list.
8	Decide	Making decisions on design ideas	• Like if you have a lot of ideas, it would be hard to choose just one. • Then he picked the best idea.

Note that in the students' responses, many referred to Chris as "he", although we had designed the task so that Chris could be considered as either a male or female student. Data collected in each school year were coded by different coders. To ensure inter-rater reliability, each coder practiced coding on at least 35 past responses and reached consensus with the researchers who developed the coding scheme.

We chose to code the responses dichotomously: each student's response received either a score of 1 or 0 in each coding category depending on whether that concept is present or not. Thus each student's overall score could range from 0-8, based on the number of categories covered in the student's response.

Data Analysis: Score distribution and comparison

A Kruskal-Wallis test was performed to compare the ranks of scores in the 2nd, 3rd, and 4th grade groups. Since significant results were found, post-hoc tests using Mann-Whitney tests with a Bonferroni correction were used to see which groups are significantly different from one another. The Bonferroni correction was used to control for Type I error.

Results: Score distribution comparison and comparison

A Kruskal-Wallis Test revealed a significant difference in the overall design process knowledge scores across the three grade groups, $\chi^2(2, n=919)=49.41$, $p<0.01$ (see Table 2 for the full set of descriptive statistics for each grade level's scores). Post-hoc tests using Mann-Whitney Tests with Bonferroni correction revealed that median scores of the three grades groups all differed significantly (Table 3). For the difference between 2nd and 4th grade groups, the effect size was medium (r=0.28), while the effect sizes for the other two group comparisons were small.

Table 2. Distribution of data for students' overall score

Grade	Mean	Median	S.D.	Upper 95%	Lower 95%
2nd (n=297)	1.22	1	1.47	1.39	1.05
3rd (n=300)	1.59	1	1.57	1.77	1.41
4th (n=322)	2.14	2	1.81	2.34	1.94

Table 3. Post-hoc Mann-Whitney Tests Results

Groups	Mann-Whitney U	Wilconxon W	Z	p	$r=Z/\sqrt{n}$
2nd/ 3rd	38121.50	823745.00	-3.17	<0.01	0.12
3rd/4th	39829.00	84979.00	-3.87	<0.01	0.15
2nd/4th	32746.00	76999.00	-6.97	<0.01	0.28

We further broke down the data by design process concepts. From Figure 2, we can observe that the 4th graders were more likely to include discussion of *Improve, Plan, Imagine* and *Test* in their responses compared to 2nd and 3rd graders, with the greatest differences between 4th and other grades was for *Improve* and *Plan*. The 3rd graders were slightly more likely to include *Improve*, *Create, Imagine, Plan* and *Test* compared to 2nd graders, with Test being the most notable difference followed by *Create* and *Improve*. Besides the differences due to grade levels, different design concepts also pose different levels of difficulty to students, with *Test* being the aspect most likely to be included in students' responses and *Ask* the least likely. Thus, we performed list analysis to characterize the difficulty of various design process concepts statistically. In Figure 2, design process concepts are arranged in order of least to mostly commented on.

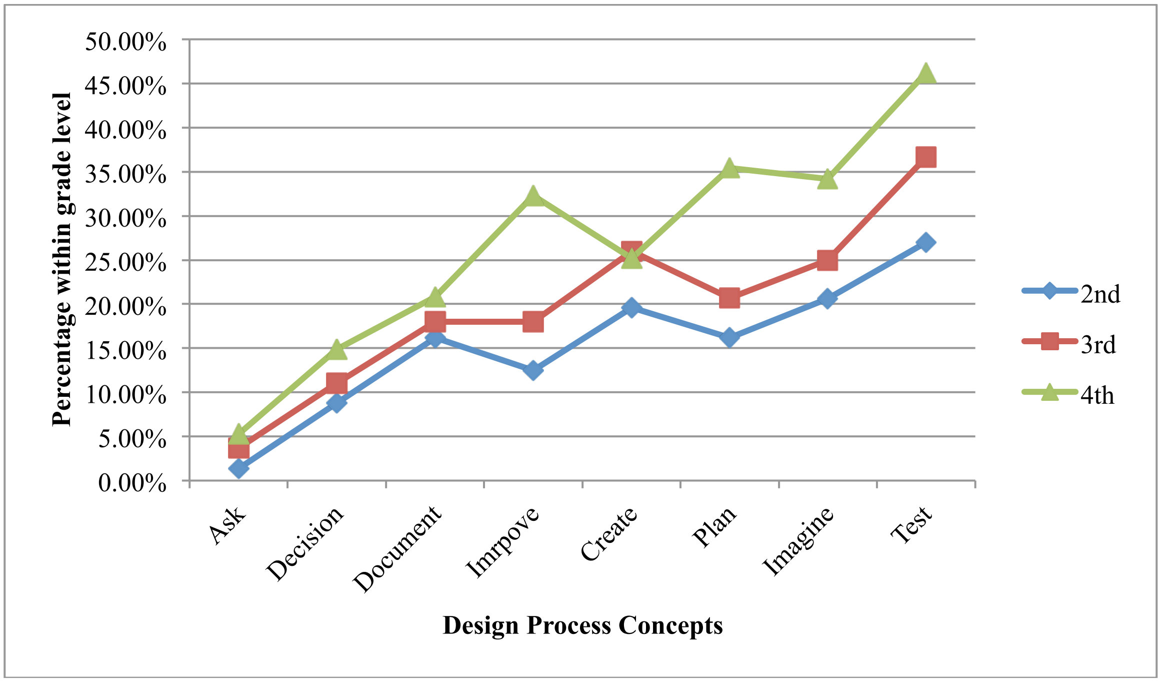

Figure 2. Percentage of students within each grade level who scored on the design process concepts

Data Analysis: List analysis

In addition to examining the learning progressions of the 2nd, 3rd and 4th graders, as measured by the design process knowledge task presented in Figure 1, we also conducted statistical analyses to better understand properties of the assessment task itself. To discern the difficulty of the design process concepts, we performed a list analysis of the student responses. In doing this, we merged the three years of data (i.e. data collected in 2009-2010 was combined with data collected in 2010-2011 and 2011-2012) but the data for each grade level was kept separate. We provide the details of the list analysis in Appendix A. As a number of stakeholders within the pre-college engineering community have recognized a need for the community to develop more valid assessment instruments, we provide the details of our process in order to provide a reference to other researchers who may wish to follow a similar process in validating their own instruments. We also provide this information to allow opportunity for the community to determine if the results we present are based on a logical chain of reasoning, or if there are areas where alternative approaches or interpretations are appropriate.

Results: List analysis

Figure 3 displays the difficulty and discrimination coefficients for each of the 8 design concepts from Table 2. For example, score1 refers to score for category 1, *Ask*; score 2 refers to score for category 2, *Imagine*, etc. Design concepts with a larger difficulty coefficient were mentioned less frequently in students' responses to the task, and design concepts with a large discrimination coefficient were more likely to differentiate a student who scored higher on the task from a student who scored lower on the task (that is, there was a pattern where students who included that particular design concept in their response overall tended to score higher on the task). We discuss this further as we discuss the patterns that emerged between and across grades.

Parameter Estimates

Item	Difficulty	Discrimination
score1	2.3216	13.1509
score2	1.0169	1.1860
score3	1.1003	1.5849
score4	1.1397	1.1209
score5	2.2755	0.8127
score6	0.7409	0.9860
score7	1.3033	1.2348
score8	5.6480	0.3930

Parameter Estimates

Item	Difficulty	Discrimination
score1	4.3509	0.7406
score2	0.4743	2.0675
score3	0.8450	1.4849
score4	0.7525	1.0926
score5	2.0384	0.6691
score6	0.3180	0.8080
score7	1.1460	1.2491
score8	3.2407	0.6088

Parameter Estimates

Item	Difficulty	Discrimination
score1	3.4616	0.8408
score2	0.4716	0.9451
score3	0.2685	1.2702
score4	0.4814	2.3634
score5	0.5079	1.0549
score6	-0.2056	1.3395
score7	1.2288	1.0164
score8	1.6352	1.0573

Figure 3. Calculations of difficulty and discrimination for grades 2, 3, and 4 from left to right. Score 1 refers to score of category 1, *Ask.*

For all grades, the difficulty of concept 1, *Ask,* is ubiquitous, but concept 8, *Decide,* provided considerable difficulty as well. With such a high degree of discrimination for grade 2, concept 1, *Ask,* may not be an ideal question when evaluating students as so few students commented on that aspect of the design process, regardless of relative difficulty. While still difficult, concept 1, *Ask,* did not differentiate between groups of students as much in grades 3 and 4, indicative of two things: first, the overall higher scores of grades 3 and 4; and second, the higher overall number of students who scored on concept 1, *Ask,* for grades 3 and 4 relative to grade 2.

In our data set, students at the high end of the ability scale in the three parameter logistical model never encountered a problem where they had a greater than 50% chance of being outperformed by those at the low end of the ability scale. That being said, some patterns emerge between our students who had high total scores and those who had lower total scores. For instance, in grade 2, our discrimination index is dominated by concept 1, *Ask*. In other words, those having a high total score have a significantly higher chance of scoring on *Ask* (i.e. including *Ask* in their response) than those with low total scores. Even though the difficulty of concept 8, *Decide* was ubiquitous, for those in grade 2, it was not a discriminatory factor in separating high scorers from low scorers.

Ideally, we want to know which aspects of design process concepts in our data set separated the high from low scorers, but did so without being overly onerous on the difficulty scale. For grade 2, concept 1, *Ask*'s high discrimination and relatively low difficulty is an example, although with such a high discrimination factor, we take the risk that very few overall students actually included this concept in their responses. A better choice would be concept 3, *Plan*, which has moderate relative difficulty and the second-highest discrimination of all of the concepts. For grade 3, concept 2, *Imagine,* obviously separated the two without a huge differential in difficulty.

For grade 4, concept 4, *Create,* and concept 6, *Test,* fit our needs. Concept 6, *Test,* is an interesting choice, as here is an example of a low difficulty (in fact, negative) score with relatively high discrimination compared to the other concepts. It may be intuitive to conclude that with such a low relative difficulty and similar discrimination index relative to say, concept 7, *Document*, that the passrate for concept6, *Test,* would be higher than concept 7, *Document* (more discussion of "passrate" is presented in step 4 in Appendix A). In fact, this is exactly the case, with the passrate of concept 6, *Test,* being nearly twice as much as that of concept 7, *Document*. So concept 6, *Test,* is a design concept that most students are able to include in their responses, and whether or not they include this concept happens to differentiate the low from high achievers in this analysis for grade 4, suggesting that the students who understand the importance of *Test*ing in the design process are more likely to recognize the importance of other design activities and include other design concepts in their responses.

Discussion

This study provides us with an understanding of how elementary students' understanding of the engineering design process progresses over the years. In general, 4th graders received higher overall scores than 3rd graders, and 3rd graders earned higher overall scores than 2nd graders, as we might anticipate. The list analysis enabled us to see the design concepts that separate the high from low scorers for each grade level: for second graders, *Plan* was the differentiating category, where high scorers tended to include discussion of *Plan* in their responses and low scorers did not. For third graders, the *Imagine* category was the main differentiator. For fourth graders, *Test* was the main differentiating category for students who scored higher on the task compared to students who scored lower on the task. In each case, it seems that the students who understood the grade-specific "gateway" design concept (*Plan, Imagine*, or *Test*) and included that concept in their response were more likely to also discuss other design concepts in their response, and thus exhibit an overall more-comprehensive understanding of the design process.

Previous studies of undergraduate engineering students' and practitioners' design behavior would suggest that it is reasonable that very few children included anything related to the *Ask* category in their responses, as the ability to engage in this behavior was a key way in which the practitioners demonstrated expertise (i.e. by engaging in far more *Ask*-like behavior than the students) [10]. However, we believe that the *Ask* category was problematic in this study in two regards. First, it did not differentiate scores; nearly all children received a score of 0 for *Ask*. Second, it may be that many students did not comment on Ask because this activity was absent from Chris' design process (see Figure 1). The intent behind the design of the originally version of the task developed by Bailey [16] was that students would be able to notice that Ask was missing, but nearly none of the elementary school students were able to accomplish this. Furthermore, in another study on the design behavior of children aged 4-11, the research team has found that in general young children are able to engage in problem scoping behaviors that align with the Ask category [27]. Therefore we can conclude that this design process knowledge task may not be effective in measuring students' ability to engage in *Ask*, though it may still be possible that children don't associate their own *Ask*-related behavior as being a part of an engineering design process. With this limitation in mind, it is difficult to draw conclusions about the level of understanding of the 2nd, 3rd and 4th graders who participated in this study with respect to the *Ask* design concept, and there is reason to consider an alternative version of this task for future studies which would include the *Ask* design concept in the prompt (i.e. include a step where Chris is engaging in *Ask* behavior).

With regards to the seven other design concepts, the data collected for this study suggests that the design process knowledge task is able to capture students' understanding of these concepts, and is able to reveal differences between 2nd vs. 3rd vs. 4th grade students' understandings of the design process. The data also provides insights into the elementary students' engineering design process learning progressions, where we see that students of different grade levels are more or less likely to include particular design concepts in their responses.

Limitation

Regarding learning progressions, the goal is to build a more sophisticated way of understanding and reasoning. Our study focused on what the difficult concepts in engineering design are at different grade levels, but not on describing what more sophisticated ways of understanding looks like. Qualitative analysis and synthesis is needed to model progressions for learning goals of engineering design. What is needed is also how understanding of separate design concepts relates to design as a whole and the pedagogical methods to achieve these learning goals.

Conclusion

The results of the study represent the current status quo of teaching and learning of engineering design, and should inspire improvement. The work presented in this paper provides insights into the learning progression of design process concepts for 2nd-4th graders, where *Plan* seems to be a key concept for 2nd graders to be able to learn, *Imagine* a key concept for 3rd graders, and *Test* a key concept for 4th graders. This may mean that 2nd grade curriculum should particularly focus on helping students develop an understanding of the *Plan* concept, 3rd grade curriculum should particularly focus on helping students develop an understanding of the *Imagine* concept and 4th grade curriculum should particularly focus on helping students develop an understanding of the *Test* concept. In each case we would envision that the process of developing an understanding of the *Plan*, *Imagine*, or *Test* design concept would include both activities focused on a cognitive or conceptual understanding but also activities that give students an opportunity to practice that design activity. Regarding the *Ask* finding, it is possible that this is a key area for all grades in terms of more explicit instruction, particularly in helping students realize that their curiosity and question asking behavior is an important part of the engineering design process. Findings from other studies examining first-year college students' *Ask* behavior [3] [6] suggest that there is indeed a need for pre-college education to help students develop these abilities. This might be accomplished through instructional activities where students are encouraged to ask questions (even an activity such as a KWL chart [19] can work to accomplish this) as well as giving students opportunities to work on open-ended problems where they are not given all of the information they need to solve the problem (in contrast to most mathematics story problems). However, we also acknowledge that the assessment task from our study may be limited in its ability to appropriately measure this concept (although the data suggests that the task is able to measure students' understanding of the other seven concepts). Future work should be conducted to better understand this phenomenon.

We hope that the research findings related to elementary students' understanding of the design process provide a foundation for future teacher professional development and curriculum development efforts, as well as research work, and that the discussion of the statistical procedures likewise provides a foundation for other researchers' work as more assessment instruments are developed for the pre-college engineering community.

Acknowledgement

This work was supported 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(s) and do not necessarily reflect the views of the National Science Foundation.

We would like to thank Dr. Heidi Diefes-Dux for leading the data collection efforts of the project and the many students at INSPIRE for assisting in data collection and transcription. We would also like to thank Delean Tolbert, Brittany Mihalec-Adkins, and Bailey Mantha-Nagrant for their help with coding the data. Finally, this work benefitted Dr. Şenay Purzer's contributions to earlier phases of this study.

References

1. ABET, *2010-2011 Criteria for Accrediting Engineering Programs*. 2009, ABET, inc: Baltimore, MD.
2. National Research Council. *Next generation science standards: For states, by states*. 2013 [cited; Available from: http://www.nextgenscience.org/
3. Atman, C., et al., *Engineering design processes: A comparison of students and expert practitioners.* Journal of Engineering Education, 2007. **96**(4): p. 359.
4. Atman, C., et al., *A comparison of freshman and senior engineering design processes: an in depth follow-up study.* Design Studies, 2005. **26**(4): p. 325-357.
5. Atman, C., et al., *A comparison of freshman and senior engineering design processes.* Design Studies, 1999. **20**(2): p. 131-152.
6. Crismond, D.P. and R.S. Adams, *The Informed Design Teaching and Learning Matrix.* Journal of Engineering Education, 2012. **101**(4): p. 738-797.
7. Adams, R.S., J. Turns, and C. Atman, *Educating effective engineering designers: the role of reflective practice.* Design Studies, 2003. **24**(3): p. 275-294.
8. Duschl, R.A., H.A. Schweingruber, and A.W. Shouse, *Taking science to school*. 2007, Washington, D.C.: National Academies Press.
9. Corcoran, T., F.A. Mosher, and A. Rogat, *Learning Progressions in Science: An Evidence-Based Approach to Reform*. 2009, Consortium for Policy Research in Education: Philadelphia, PA. p. 86.
10. Schwartz, D.L., S. Varma, and L. Martin, *Dynamic transfer and innovation*, in *International Handbook of Research on Conceptual Change*, S. Vosniadou, Editor. 2008, Routledge: London, UK.
11. Duncan, R.G. and C.E. Hmelo-Silver, *Learning progressions: Aligning curriculum, instruction, and assessment.* Journal of Research in Science Teaching, 2009. **46**(6): p. 606-609.
12. Mohan, L., J. Chen, and C.W. Anderson, *Developing a multi-year learning progression for carbon cycling in socio-ecological systems.* Journal of Research in Science Teaching, 2009. **46**(6): p. 675-698.
13. Songer, N.B., B. Kelcey, and A.W. Gotwals, *How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity.* Journal of Research in Science Teaching, 2009. **46**(6): p. 610-631.
14. Schwarz, C.V., et al., *Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners.* Journal of Research in Science Teaching, 2009. **46**(6): p. 632-654.
15. National Committee on Science Education Standards and Assessment and National Research Council, *National Science Standards*. 1996, Washington, DC: The National Academies Press.
16. Bailey, R. and Z. Szabo, *Assessing engineering design process knowledge.* International Journal of Engineering Education, 2006. **22**(3): p. 508-518.

17. Hsu, M.-C., M.E. Cardella, and S. Purzer. *Elementary Students' Engineering Design Process Knowledge: Instrument Development and Pilot Test*. in *American Society for Engineering Education Annual Conference*. 2012. St. Antonio, TX.
18. Cunningham, C. and C. Lachapelle. *The impact of Engineering is Elementary (EiE) on students' attitudes toward engineering and science*. in *American Society for Engineering Education Annual Conference & Exposition*. 2010. Louisville, KY.
19. Ogle, D.M., *KWL: A teaching model that develops active reading of expository text.* Reading teacher, 1986.
20. Baker, F., *The Basics of Item Response Theory*. 2001, University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
21. Glas, C.A., *Item Parameter Estimation and Item Fit Analysis*, in *Elements of Adaptive Testing*, W.J. van der Linden and C.A. Glas, Editors. 2010, Springer New York. p. 269-288.
22. Codey, R.P., *Psychometrics 101*. 1999, Robert Wood Johnson Medical School.
23. SAS Institute, I. *The Item Analysis Report*. 2014 [cited; Available from: http://www.jmp.com/support/help/The_Item_Analysis_Report.shtml.
24. SAS Institute, I. *Computing Confidence Limits for Quantiles and Percentiles*. 2014 [cited; Available from: http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_univariate_sect065.htm.
25. SAS Institute, I. *Estimating Three Parameters from Lognormal Quantile Plots*. 2014 [cited; Available from: http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_univariate_sect086.htm.
26. Beam, D., *SUM Enchanted Procedure: The Advanced Power of PROC Summary*. 2011, Systems Seminar Consultants. p. 1-6.
27. Ricco, G., et al. *Sectionality or Why Section Determines Grades: an Exploration of Engineering Core Course Section Grades using a Hierarchical Linear Model and the Multiple-Institution Database for Investigating Engineering Longitudinal Development*. in *American Society for Engineering Education Annual Conference*. 2012.
28. Chong, H.Y. and J. Wai-Chi Wong, *Using SAS for Classical Item Analysis and Option Analysis*, Arizona State University.

Appendix A: Extended Discussion of the List Analysis

To discern the difficulty of the design process concepts, we performed list analysis in SAS to student responses with the three years of data collection merged. Moreover, to explore the possible differences, the analysis was performed by separating data based on grade level.

List or item analysis refers to a large body of interpretive (or descriptive) statistics throughout multiple fields, including statistics, physics, economics, biology, education, sociology, and more. The purpose of this analysis is to provide not only a basic item analysis on the data set at hand, but a short primer on its values. For instance, some of the caveats of item analysis are not exactly intuitive. As an example, when first learning about score distribution plots and their relationship to univariate analysis routines, one may be surprised to learn that the assumption of *normality* does and does not matter! In other words, it is of extreme importance to understand one's data, its relationship to the analysis method chosen, and even more importantly, the exceptions and strange quips involved.

First of all, for the interested, there is no end of robust free resources available for list analysis on the Internet and in hard copy. For those familiar with SAS and SAS JMP, we include some links to example codes, but all of them contain great examples and interpretations. Here are some of our favourites that are easily accessible: Frank Baker's entire book on the Basics of Item Response Theory, which will provide not only an explanation of all of the methods employed here, but their mathematical underpinnings [20]; Glas's book (in a series) on item parameter estimation and item fit analysis, which contains a technical description of nearly every method a test evaluating academic may hope to employ [21]; an introductory primer from a psychometrics course by Robert Codey [22]; the SAS Institute's numerous pages which contain not only example SAS code and SAS JMP instructions, but contain example data, interpretations, and links [23-25]; and great article by Beam on the most fundamental item analysis procedure in SAS, the PROC SUM [26].

As with any survey analysis involving multiple questions, we employed a general algorithm like the following as the procedure to interpret list results:

1. Calculate individual student scores (calibrated against any post-examination manipulation, such as grade adjustments and instructor disposal of "bad" questions).

2. Produce basic statistics (mean, median, mode, determine type of distribution of scores themselves).

3. Plot histograms of relevant total grades and question distributions. Once again, there is nothing mathematically complicated about this part of the analysis. Figure 4 is an example of such a plot for our data, broken down by student grade, where each "question" in this case is one design concept that a student might include in their response.

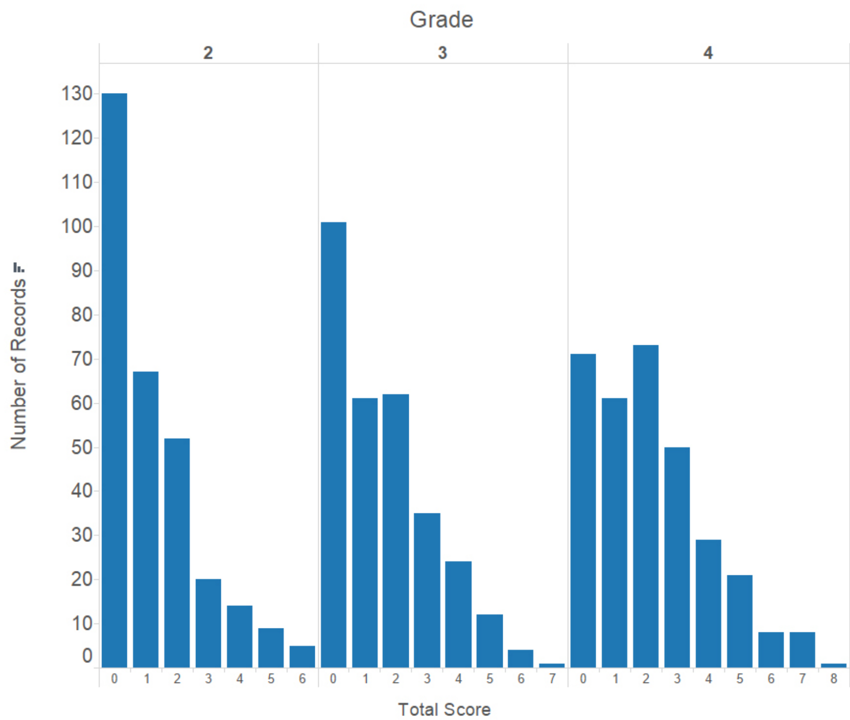

Figure 4. Histogram of total scores broken down by grade

4. Produce a question ranking for each individual question (or in this case, design concept). This is a number that ranks the percent *right* or *wrong* relative to either an empirically *right* answer, an expert's answer, or some combination of factors. In SAS, this is given by a *procedure summary* call when we have our data arranged in one-dimension form, as demonstrated in Table 4. If the data is not in one-dimensional form, SAS can obtain this data through a transformation. After we place the data in this form, we can run the summary for long form data or choose the *procedure means* call. Table 5 is an example of the pass rate for the entire set of data analyzed here. Ignore the analysis of the *null* represented by a period and a type of "0," as it is insignificant here.

Table 4. Example of *one dimensional* data (*Number* here refers to the coding category number and *response* here refers to the student's response for that design concept)

Number	Response
1	0
2	0
3	0
4	0
5	0
6	1
7	1
8	1
1	0
2	1
3	0
4	1
5	1
6	0
7	1
8	0

Table 5. Example of a *passrate* analysis from a SAS proc summary call

Number	_TYPE_	_FREQ_	passrate
.	0	7352	0.2080272109
1	1	919	0.0348583878
2	1	919	0.2676822633
3	1	919	0.2437431991
4	1	919	0.2361262242
5	1	919	0.2124183007
6	1	919	0.3688792165
7	1	919	0.1838955386
8	1	919	0.1164309032

5. Use a quantile routine to *rank* student total grades. This can be done one of two ways: a simple counting quantile routine, as is available in Excel and other spreadsheet packages; second, a mathematical quantile ranking (or quantile function based approach,) which involves the rank of grades based on a univariate analysis routine easily performed in SAS, SPSS, with Excel add-ons, calculated by hand, etc.

   Quantile function analysis itself is a tricky beast, even if one is using a standard procedure. In Excel, quantile calculations can be done with add-ons and VBA scripts easily available on the web. In SAS, SAS JMP, and SPSS, there are multiple ways to produce a quantile plot, but all involve some invocation of a univariate procedure. When ranking in SAS, the basic ranking procedure assumes a normal distribution (aka it uses a normal quantile function distribution to fit,) but it does *not* matter if the data distribution itself looks very *normal*. There are multiple reason for this, the first and foremost is that the quantile rank itself is a useful diagnostic tool that will allow the researcher to determine whether or not a logarithmic or other fit is useful for the student data. Depending on the nature of the research and the exam itself, one may choose to use the quantile as an indicator that a Q-Q plot should be used next. Quantile distribution functions themselves can take many forms, for example, the exponential quantile takes the form,

$$Q(x, y) = -\frac{\log_e(1-x)}{y}$$

   Where, x is greater than 0 and less than 1.

   Using a simple quantile ranking procedure in SAS to rank quantiles via a counting procedure, we obtain Table 6. Not included in Table 5 is the ranking of students, i.e., the number of students below and above each quantile.

Table 6. Basic quantiles for grades 2, 3, and 4 going left to right

Quantile	Estimate	90% Confidence Limits Distribution Free		Quantile	Estimate	90% Confidence Limits Distribution Free		Quantile	Estimate	90% Confidence Limits Distribution Free	
100% Max	6			100% Max	7			100% Max	8		
99%	6	5	6	99%	6	5	7	99%	7	7	8
95%	4	4	5	95%	5	4	5	95%	6	5	6
90%	3	3	4	90%	4	4	4	90%	5	4	5
75% Q3	2	2	2	75% Q3	3	2	3	75% Q3	3	3	3
50% Median	1	1	1	50% Median	1	1	2	50% Median	2	2	2
25% Q1	0	0	0	25% Q1	0	0	0	25% Q1	1	0	1
10%	0	0	0	10%	0	0	0	10%	0	0	0
5%	0	0	0	5%	0	0	0	5%	0	0	0
1%	0	0	0	1%	0	0	0	1%	0	0	0
0% Min	0			0% Min	0			0% Min	0		

6. From the results of the mathematical quantile ranking, one can easily determine if the distribution of ranked grades is normal, logarithmic, or some other type of distribution. From Table 5, we cannot easily tell if our distribution is better fit by a logarithmic quantile function, that is because it is a simple numeric count based on an *ordinal* distribution. Examining Figure 4, it is somewhat apparent that the distribution of total grades for grade 2 is not exactly normal. If we were to use a more advanced quantile function fit in SAS, the resulting ranks would yield negative values where the zeroes in Table 4 are right now. That is a typical example of when a quantile function fit is a diagnostic tool to determine whether or not further testing is required.

7. In the case that a quantile function returns an interesting (aka not normal) distribution, a Q-Q plot may be a good idea. The quantiles in a Q-Q plot are, by definition log normal. This means that in the case of our total distribution for grade 2 (See Figure 5), where a quantile function fit would yield negative values, a Q-Q logarithmic plot of the data will appear to be a fairly straight line (except for areas of the curve where the number of student scores is extremely low, such as when total = 6).

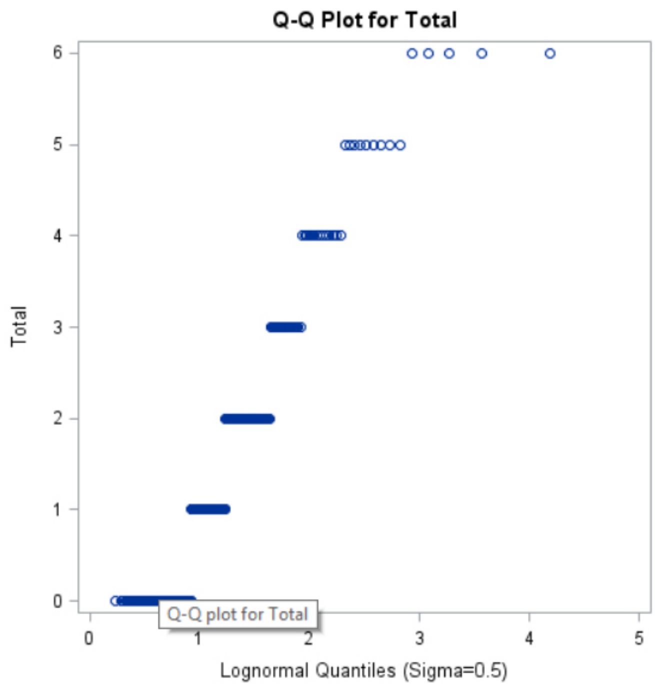

Figure 5. A Q-Q plot for our data of grade 2.

If we were to draw a line through the plot here, we would discover that due to the weighting of the distribution, even including the values at Total = 6, that we would have a straight line with less deviation than, say, a Q-Q plot of our relatively normal distributions found in grades 3, 4, and a plot of all grades combined.

8. Here is where things become interesting. From here, we have a choice of deriving biserial and point biserial calculations, which are roughly variations of inner product spaces. The point biserial calculation (or point biserial correlation coefficient) in item analysis is useful for comparing the relative difficulty of one problem relative to the test itself. The point biserial correlation coefficient follows the same basic rules as the intraclass correlation coefficient. Brief primers for calculating the point biserial correlation coefficient and the intraclass correlation coefficient are available elsewhere [27, 28]. We have forgone calculating the point biserial correlation coefficient here due to calculating the indices of discrimination and difficulty.

9. Calculate the index of difficulty and the index of discrimination for each problem in a given group for the test items. The index of difficulty is a measurement of the number of students who guessed the wrong answer on a problem relative to the entire group. The index of discrimination is a relative measurement between any group of high ranked students and any group of low rank students. The parameterization of *high* and *low* can be based on quantiles, *information* and *ability* curve (classic list analysis does this,) or any other discriminating factor the researcher chooses. Technically, the high and low groupings of students could be based on *expert* answers, which makes its construction and calculation flexible.

As a side note, depending on how the two coefficients are calculated, some interesting phenomena can occur. For instance, the higher the difficulty index usually indicates an *easier* problem, yet many software packages (SAS, SAS JMP, SPSS, and more,) often normalize the difficulty coffecient depending on the routine so that a higher value does not mean it is easier to answer, but in fact the exact inverse. Another interesting example is that as the discrimination index becomes *negative,* this indicates that for students whose overall test score is *high*, a certain problem presented unique hurdles to overcome.