Validation of Approaches to Assess Design Process Knowledge

Reid Bailey and Zsuzsanna Szabo
College of Engineering/Educational Psychology
University of Arizona

ABSTRACT

Rigorously assessing students' design process knowledge is essential for understanding how to best create learning environments to facilitate the development of such knowledge. Such assessment is also quite difficult and hence, no assessment tool capable of measuring design process knowledge of every student in a large college exists. Faculty from both the Colleges of Engineering and Education at the University of Arizona are developing such a tool. In this paper, results from the first year of implementation of the design process knowledge assessment tool are presented. The goal of the first year was to collect and analyze data that can be used to validate and improve the tool. Results from such analysis, as well as an overview of the tool itself, are presented in this paper.

MOTIVATION

A core learning objective for engineering students from all disciplines at all universities is to learn about engineering design. To this end, capstone design courses populate nearly all curricula while design courses in freshman and other years are becoming more commonplace. Despite the ubiquity of engineering design in curricula, little if anything is known about what students learn in engineering design courses. The authors seek to develop a tool to remedy this lack of knowledge. In this paper, results from the first round of validation of this tool are presented.

CONTEXT

A process of engineering design is subjective in that there are no mathematical proofs or conclusive experiments to prove that one process is *the* process. That said, some common elements of engineering design have emerged over the course of centuries of engineering. These common elements are seen today throughout the disciplines of engineering in education and in practice (albeit in varying forms). Engineers 1) clarify and articulate a need, 2) create a design to meet that need, and 3) implement that design. These three phases of design are typically iterated through several times before a design is finalized. This process is shown in Figure 1.

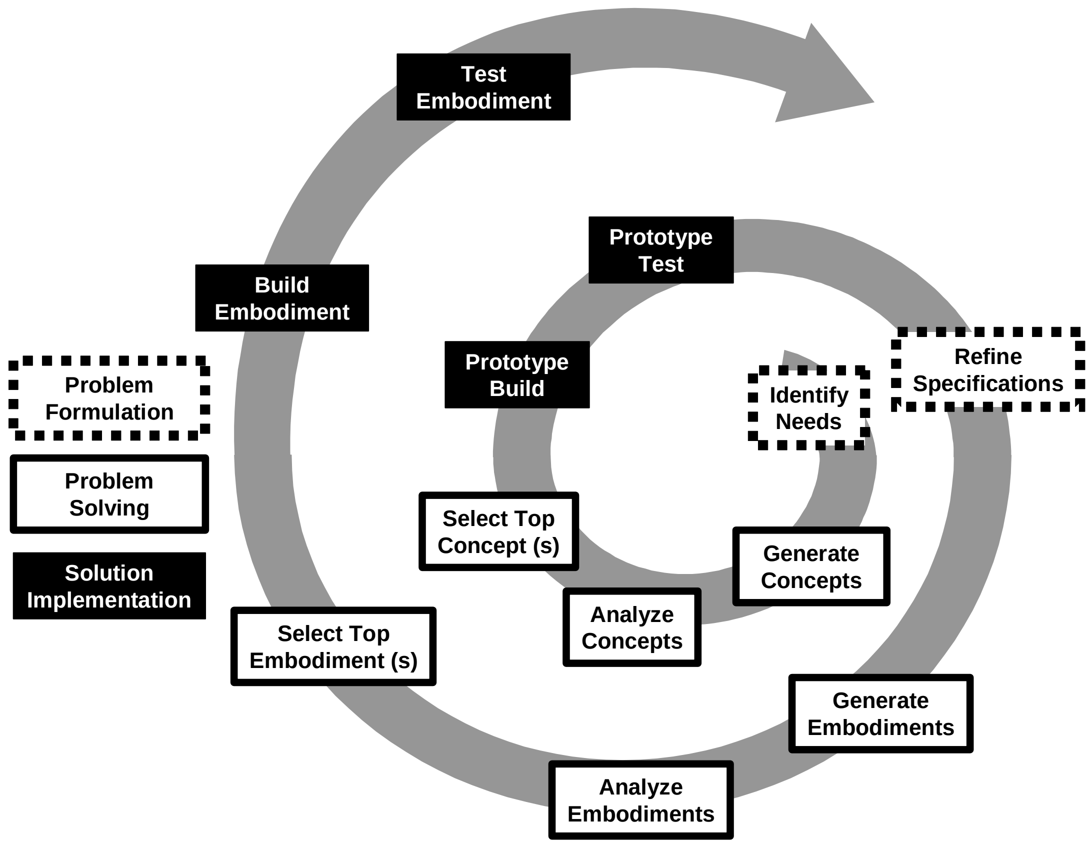

Figure 1 Engineering Design Spiral

The process in Figure 1 starts at the center of the spiral. "Problem Formulation" steps relate to identifying and clarifying the needs to be met by the engineers. "Problem Solving" steps involve developing designs on paper. Problem Solving includes the divergent process of creating several solutions to a problem and the convergent steps of analyzing this set of solutions and selecting the more promising solutions to implement in the third phase. "Solution Implementation" is focused on turning ideas on paper into realized systems. The two primary activities within Solution Implementation are building[*] the design and testing it. The spiral is used to represent iteration through these three phases. While each iteration may not include every step of each phase, each of the phases is found in nearly any engineering design process at some point.

The work presented in this paper is aimed at assessing the following overall instructional objective: students should be able to explain and analyze a design process involving iteration through the three phases shown in Figure 1. The target group includes both first year engineering students enrolled in an introduction to engineering design course and seniors in capstone design courses. This instructional objective is linked to multiple levels from Bloom's taxonomy, a set of six basic types of cognitive learning[4]. A revised version of Bloom's taxonomy contains the following six levels[1,7]:

---

[*] The word "building" is used broadly to include not only physically building a system, but also activities such as writing software code.

1. *Remembering*: Being able to recite information from memory without necessarily understanding it.
2. *Understanding:* Being able to explain material within its own domain.
3. *Applying*: Being able to use a concept to solve a particular problem.
4. *Analyzing:* Being able to parse something into its parts.
5. *Evaluating*: Being able to judge different concepts and determine their value.
6. *Creating:* Combining concepts to create something new.

These six levels of the revised Bloom's taxonomy are related to the engineering design model from Figure 1 in Table 1.

Table 1 Engineering Design Learning Related to the Revised Bloom's Taxonomy

Remembering	Redrawing the spiral figure in Figure 1.
Understanding	Explaining Figure 1 and each phase represented in it.
Applying	Implementing the process depicted in Figure 1.
Analyzing	A higher level of understanding and application where the purpose of each step is clearly understood and only used when necessary.
Evaluating	Comparing the process in Figure 1 to other design processes and explaining the strengths and weaknesses of each process.
Creating	Forming an entirely new design process.

It is expected that students in the introduction to engineering design class can remember, understand, and apply the engineering design process in Figure 1. For seniors, students should also be able to analyze the steps being used and begin to evaluate the effectiveness of alternative design processes (i.e., different specific manifestations of the process in Figure 1).

As will be explained in the following section, the approach investigated in this paper is aimed at assessing if students can explain (Understand) and Analyze an engineering design process by having them critique (Evaluate) a proposed process in the context of what they learned in class. Students are asked to compare the process in Figure 1 to a very poor proposed process; such a comparison is more akin to explaining the purpose of each step of design than it is to a serious evaluation of two similarly strong processes.

Assessment of Engineering Design Process Knowledge

In creating an assessment strategy for engineering design process knowledge, the following are key criteria. The strategy must be:

- At the individual, not team, level
- Process-focused (not only focused on quality of end result)
- Not too time-intensive (not requiring significant class time or unreasonable amounts of time to prepare and score)

- Reliable from student to student, project to project, and year to year
- Linked to more than just one level on Bloom's taxonomy

A trade study of four basic assessment strategy options in shown in Table 2.

Table 2 Options for Assessing Engineering Design Process Knowledge

Assessment Option	Positives	Negatives	Scoring Method
Base assessment on **Design Reports** that students already turn in as part of class	+ Process-focused + Already a part of class, so not too time intensive.	– Not at individual level – Potential Interrater reliability problem if multiple graders – Only linked to application level of Bloom's taxonomy	• Checklist rating scale
Base assessment on the **Performance of Final Designs**	+ "objective"/measurable + Already a part of class, so not too time intensive.	– Not process focused – Too specific to project – Not at individual level – Potential interrater reliability problem if multiple graders – Only linked to application level of Bloom's taxonomy.	• Rubric-based assessment of final designs.
Base assessment on responses to **Questions** (multiple choice/short answer/essay)	+ At individual level + Can be process-focused + Can be linked to any level of Bloom's taxonomy	– Potential interrater reliability problem if multiple graders – Can be time intensive. – Can questions really assess design skills?	• Analytic or holistic rubric-based assessment of answers to questions.
Base assessment on **Video** of design teams "in action" or reflecting on process	+ Process-focused + Can be at individual and team levels	– Time intensive for 800+ students per year! – Only linked to application level of Bloom's taxonomy.	• Performance assessment.

It is clear that each approach in Table 2 has strengths and weaknesses. Not being at the individual level is a big weakness of both design reports and final designs, with final designs also being hampered because they are not process-focused. Using video would require a prohibitive amount of time to watch and reliably score the tapes. The remaining option, having students respond to questions, can be process-focused and at an individual level. The major weaknesses of this approach are potential interrater reliability problems, time intensives, and some doubt as to whether it can truly measure design skills. Problems with interrater reliability occur when multiple people score the same response differently. Each of these weaknesses can be addressed with a well-designed and validated tool, whereas the weaknesses of the other options are much more difficult or impossible to address. Hence, a question-based assessment strategy is used.

The fourth column in Table 2 indicates the scoring method associated with each assessment option. In each case, a well-constructed rubric is needed to generate reliable scores for the student deliverables. In the following section, a review of rubrics is presented.

Assessment of Student Learning

Authentic assessment (performance assessment) is used when the goal of assessment is to observe the process of thinking or the actual behavior of a performance task. The authenticity of a test is measured by how closely the test assesses the knowledge and skills required in real life[5]. The essence of authentic assessment is performance. Performance assessment requires students to use their knowledge and produce something (e.g., a group project, a demonstration, to build something, produce a report). As defined by Nitko[6], performance assessment requires students to apply their knowledge and skills and presents them with a hands-on task that requires them to do an activity.

In order to evaluate how well the students have achieved the task, clearly defined criteria are used. When performance assessment is used, students are required to demonstrate their achievement by producing a developed written or spoken answer that will demonstrate their achievement of a learning target. The performance task can be used to assess the process, the product, or both. To assess the performance task, scoring rubrics or rating scales are used.

The use of essay responses in student assessment[6], on the other hand, permits the measurement of students' ability to describe relationships, apply principles, present relevant arguments, state necessary assumptions, describe limitations of data, explain methods and procedures, produce, organize, and express ideas, evaluate the worth of ideas, etc. The use of essay assessment is beneficial in assessing higher-order thinking skills. For this reason, short essay responses were used in assessing engineering students' learning in project-based courses. When using essay assessment, it is very important to set well-defined criteria describing how the essays will be graded. Nitko[6] mentions two general methods for scoring essays: the analytic method and the holistic method. A top-down method is used for crafting an analytic rubric. An analytic scoring rubric requires first an outline containing a list of ideal points, major traits, and elements that a student should include in an ideal answer. The teacher would decide the number of points awarded for each element in the ideal response. Students who respond correctly to that element get the full credit, as compared to those who responded incorrectly and receive no points for that element.

The holistic rubric[6] assesses an overall impression of the response in a less objective manner than the analytic rubric. In crafting a holistic rubric, a teacher would use a bottom-up method. In the bottom-up method, the teacher begins using actual student responses of different qualities and sorting the responses in categories that would help identify the different levels of students responses. After students' responses are sorted, the teacher writes very specific reasons why each of the responses was put in the respective category. Then, for each category, the teacher writes a specific student-centered description of the expected response at that level. These descriptions constitute the scoring rubric to grade new responses.

The two methods (analytic and holistic) are not interchangeable, and the clear advantage of the analytic rubric is that it provides a more objective way of assessing students' strengths and weaknesses. Also, the analytic rubric can give teachers a clearer look over the elements where students have difficulties in answering, and might need to be retaught. The disadvantage of using analytic rubrics for assessment compared to holistic rubric is that student performances are compared to ready-set standards developed in accordance to teachers' expectations about what students are supposed to know. The ideal answer might not always reflect what students really would be able to answer based on what was taught. Hence, analytic rubrics usually undergo many revisions as the top-down expectations are adjusted to better match actual student responses.

The scoring in both holistic and analytic rubrics is slower than when objective items are used (e.g., true-false, multiple choice, matching). In the assessment presented in this paper, the analytic rubric with a top-down crafting method (and many revisions) was selected due to its increased objectivity and ability to target specific elements where students are excelling or having trouble.

STRATEGY FOR VALIDATION

The purpose of the work presented in this paper is to validate that the developed questions and analytic rubric reliably measure students' design process knowledge. There is one question in which students critique a process to design a shopping cart and another where they critique a process to design an egg counter for eggs traveling down a conveyor belt. One master rubric was developed and then particularized to the two versions of the question.

The two questions were given as a pretest and posttest pair to students in the introduction to engineering design class in both Fall 2003 and Spring 2004 (total population size of roughly 300 students)[2]. Additionally, seniors in two different capstone classes completed both the shopping cart and egg counter tests back-to-back near the end of their two-semester classes (total population of 104 students). Some of the seniors took the shopping cart test first and others the egg counter first. Between these two sets of data – the introduction class and the capstone class – the validity of each question and of the analytic rubric is investigated.

Pre and Post Tests

The shopping cart and egg counter tests are shown in Figures 2 and 3. In each case, students are instructed to identify the strengths and weaknesses of the proposed process and to explain why something is a strength or weakness.

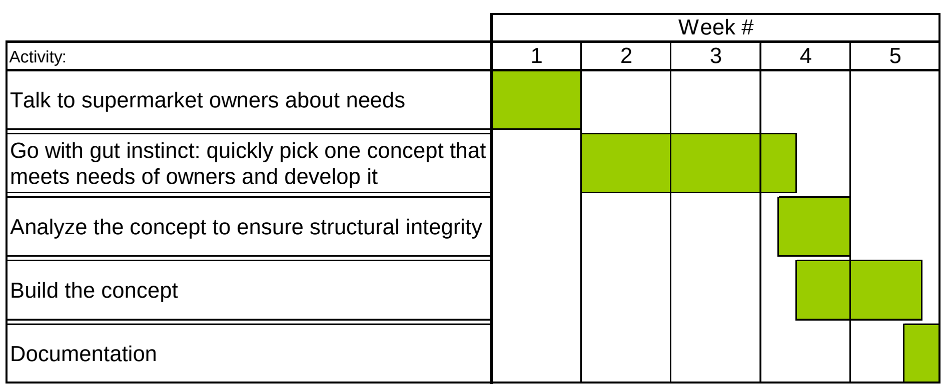

Figure 2 Shopping Cart Question

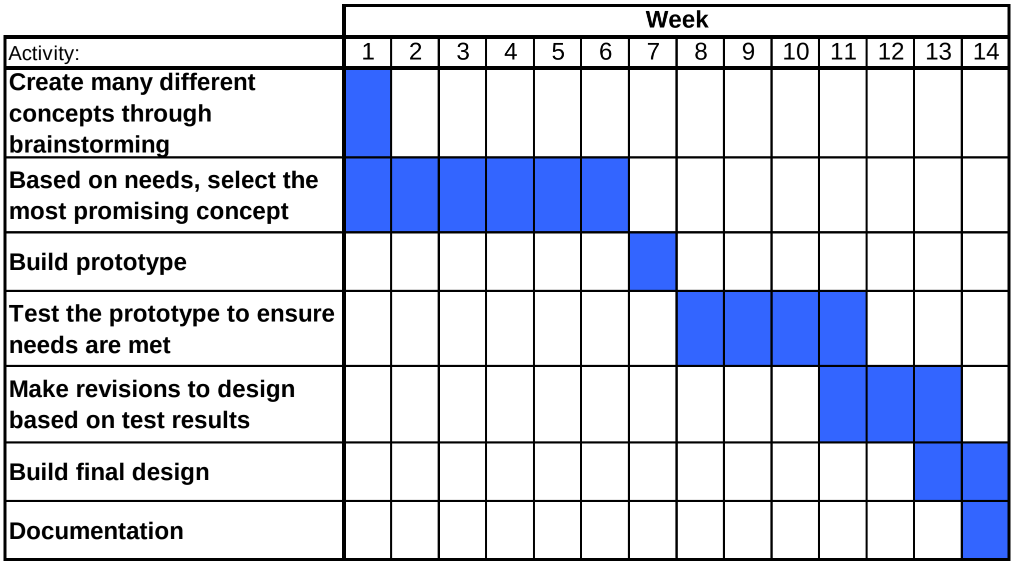

Figure 3 Egg Counter Question

The two tests are purposefully developed to not have the same strengths and weaknesses. For instance, the process proposed for the shopping cart does not do a good job of generating several alternatives before selecting one to develop whereas the egg counter does do this well. As will be discussed in a later section, however, such differences led to difficulties in validating the questions.

Rubric

The rubric for each question is derived from a master rubric that is based on the common elements of design shown in Figure 1. The rubrics are split into seventeen different levels, each focused on a different aspect of engineering design and a specific instructional objective. The shopping cart rubric is in the appendix of this paper.

Within the overall objective of “students should be able to explain and analyze a design process in the context of the three phases shown in Figure 1,” there are several more specific sub-objectives measured by the rubric. These objectives are as follows – students should be able to:

1. Explain that engineers work in teams to design systems.
2. Explain why needs must be gathered and analyze the effectiveness of techniques for gathering needs.
3. Explain why multiple alternatives should be generated before developing a single alternative in depth.
4. Explain that a combination of analysis and decision-making (based on the needs of the project) is required to eliminate ideas before building them and analyze the completeness of approaches used in analysis and decision-making.
5. Explain that built designs should be tested to determine if they meet the needs.

6. Explain how the three phases of design fit together and involve iteration, and analyze how much time is necessary for each step.
7. Explain that documentation must occur throughout a design process.

Note how each instructional objective involves *explaining* and/or *analysis* – clearly tying these objectives to the second and fourth levels of Bloom's taxonomy. The nature of the questions is more strongly tied to the evaluation level on Bloom's taxonomy (asking students to critique a proposed design process). In this assessment, however, evaluation is used primarily as a means to elicit *explanations* and *analysis* of an engineering design process taught in class (which includes the common elements in the process shown in Figure 1). To demonstrate how the rubric assesses these instructional objectives, a level from the rubric is presented in Figure 4.

Design Phase	Step	Pts.	Description	Shopping Cart
II	4	3	<u>Analyze</u> ideas on <u>all</u> relevant criteria and constraints Possible means of analysis include (do not have to mention any of these, but these are key words to look for for analysis): • Experiments/Design of Experiments • Equations/Analytical Models • Simulation • Verbal analysis through group discussion of designs	Positive: Analyzed concept for structural integrity, or, indicates that more time is needed for analysis of structural integrity (+1 pt) Negative: other areas besides structural integrity (e.g., weight, steering, ergonomics) need to be analyzed +1 pt for noting that more analysis is needed +1 pt. for noting an additional type of analysis needed

Refers to Phase from Figure 1.

Unique Step I.D. to Identify this Level.

Total # of points for this level.

General description of this level. This column is the same for all pre and posttests.

Description and point allocation particularized for the shopping cart. "Positive" means that something is done well in the proposed process in Figure 2. "Negative" means that something is omitted or needs to be changed in the proposed design process

Figure 4 Analysis Level of Rubric[2]

In Figure 4, the master rubric is in the column labeled "Description" while the shopping cart scoring is in the column labeled "Shopping Cart." A response that indicates that doing analysis before building the design is good receives 1 point. If the response also indicates that other analyses, such as mass or ergonomic analyses, are needed, then they would receive all 3 points for this level of the rubric. This directly relates to the fourth instructional objective in the preceding list. Additional examples of how to score student responses with this rubric are shown by Bailey, et al.[2].

RESULTS AND DISCUSSION

<u>*Statistical Analysis of the Data*</u>

Results from the statistical analysis of data collected in Fall 2003 and Spring 2004 seem to show that students in the introduction to engineering class learn statistically significant design process content across the semester (Fall 2003: $t = 5.14$, df=178, $p<0.0001$; Spring 2004: $t = 4.77$, df = 125, $p<0.0001$). With these students, the shopping cart was used as a pretest and the egg counter was used as a posttest. When both tests are given to seniors in one sitting, however, the average score for the egg counter was higher than that for the shopping cart. Furthermore, when these two tests were given in different orders to senior students, the results show that senior students obtained statistically significant different scores for the shopping cart ($t= 2.38$; $df=104$, $p=0.019$) but not statistically significant different scores for the egg counter ($t=0.17$, $df=104$, $p=0.862$). This means that the order in which the tests are taken is important for the shopping cart but not for the egg counter. The results indicate that the two tests are not as parallel as intended and, consequently, assess different objectives.

Item correlation analysis was conducted and the results show that in both tests (shopping cart and egg counter) there are some items that do not correlate with the overall score. There were five common items that did not work in either rubric. These five items are as follows:

- Item 1: State that a team is needed to work on the project.
- Item A: Indicate that the three phases of design are addressed in an appropriate order.
- Item C: State that iteration should be planned into a design process.
- Item H: Extra points for indicating a strength of the proposed process not listed elsewhere in the rubric.
- Item I: Negative points for answers that directly oppose to a correct answer.

The split-half correlation coefficient (between the two tests) for students in the introduction to engineering class was 0.255, and for senior students the split-half correlation coefficient was 0.497. The results show that the two tests work better for the senior students (this is expected since senior students are supposed to know more content about engineering design). The common items that are not reliable show that the rubric is not well designed and that some items would work better if they were collapsed with similar items. The rubric with seventeen items is too detailed for grading students' short answer responses (10-15 minutes response time). Even though it is possible to train raters that are not familiar with the engineering field to use the analytic rubric[3], the training is longer and needs many exercises in order to improve interrater reliability. A shorter rubric would be more advantageous in training the raters and would also decrease grading time.

In conclusion, there are two main problems identified with the statistical analysis. First, the two tests – the egg counter and the shopping cart – measure different things. This is shown with the data from seniors who took the two tests back-to-back. Second, the rubric needs to be redesigned and the rubric items improved. This is indicated with data from both sets of students using item correlation analysis.

Plans to Correct Identified Problems

To address the problem between the two tests, the shopping cart test is going to be changed to be more parallel to the egg counter test. Whereas the current tests show processes with different strengths and weaknesses, the new shopping cart test will have the same strengths and weaknesses as the egg counter test. This change should greatly increase the degree to which the two tests measure the same content.

To address problems with the rubric, the seventeen levels of the rubric will be collapsed into 4-5 levels. A new rubric with 4-5 items would be sufficient to measure the short answers provided by students and would be more efficient in training raters and in grading. Several of the problem areas on the rubric (Items 1, A, H, and I) will be removed, whereas other parts of the rubric will be joined to reduce the total number of levels.

Additionally, a second assessment tool is being developed to provide increased validity to the measurements. This tool involves students self-reporting on the design process they used in an in-class design project and also describing the process they would use to complete the in-class project (in class, teams develop concepts but do not build any prototypes or do any more detailed design work). This tool is similar to concept mapping in that students must graphically represent all the steps they use to take a project from needs to a final product. The open-endedness of this tool has proven to generate a wide range of responses from students – which certainly will be a challenge to reliably score. Rubric development for this tool will take place in Spring 2005.

CLOSURE

Because design process knowledge is less concrete than most of engineering, assessing if students are learning it is very difficult. The first round of development of an assessment strategy has been completed and statistical analysis indicates that changes are necessary to increase the validity of the tool. The analysis gives clear direction with respect to areas that need adjustment. The questions asked to the students need to be more parallel in structure and the rubric used to score responses needs to be simpler. These changes will be implemented in Spring and Fall semesters of 2005.

APPENDIX

The shopping cart rubric is presented in Tables A.1 and A.2.

Table A.1 First Eight Levels of Shopping Cart Rubric

Design Phase	Step	Pts.	Description	Shopping Cart
I	1	0.5	State that a team must be formed for the project.	Negative: +0.5 pts if stated that a team is needed
I	2	4 < 4 earned if this step is addressed but • multiple sources are not addressed • sources are not comprehensive	Gather information about project needs from multiple sources: sources should include: • All users (current and potential) of this type of device (e.g., shoppers, store owners, children) • Library and on-line research (e.g., information on injury statistics associated with shopping carts) • Existing designs – from literature (e.g., information from current manufacturers)and from direct use of existing designs (e.g., using a standard shopping cart) Information gathered is used to form criteria and constraints for the project.	Positive: information is gathered about needs (+1.5 pts) Negative: Only one source used to gather information (shop owners): +1.5 for noting that more sources are needed +0.5 for noting one additional source +0.5 for noting 2 or more additional sources Additional sources include: ➢ customers ➢ baggers ➢ research on injuries ➢ research on existing products ➢ children ➢ using the cart themselves
II	3	3	Generate multiple ideas to address the project needs through brainstorming	Negative: +3 pts if stated that they need to develop more than just one idea
II	4	3	Analyze ideas on all relevant criteria and constraints Possible means of analysis include (do not have to mention any of these, but these are key words to look for for analysis): • Experiments/Design of Experiments • Equations/Analytical Models • Simulation • Verbal analysis through group discussion of designs	Positive: Analyzed concept for structural integrity, or, indicates that more time is needed for analysis of structural integrity (+1 pt) Negative: other areas besides structural integrity (e.g., weight, steering, ergonomics) need to be analyzed +1 pt for noting that more analysis is needed +1 pt. for noting an additional type of analysis needed
II	5	2	Based on the analysis, decide which idea best meets the criteria without violating any constraints (may retain more than one concept if further iterations eventually reduce it to one final concept) Decision-making may include (do not have to mention any of these, but these are key words to look for for decision-making): • Voting • Selecting concept that maximizes a single objective • Reaching group consensus • Using a decision tool	Negative: They plan to go with "gut instinct" to choose which design to move forward with. +1 pt for stating that going with "gut instinct" is not good practice +0.5 pts for stating an alternative to going with gut instinct, such as voting, weighing strengths and weaknesses of multiple designs, considering multiple objectives +0.5 pts for stating that you should analyze your design before making decisions (i.e, before "going with gut instinct"
III	6	--	Plan how to build the selected concept	N/A for shopping cart
III	7	3	Build the concept	Positive: The concept was built: this must be directly addressed to get 3 pts.
III	8	3	Test the built concept to determine how well criteria and constraints are met	Negative: The built cart is never tested. +3 pts for stating this.

Table A.2 Final Nine Levels of the Shopping Cart Rubric

	Pts	Description	Shopping Cart
A	3	The 3 phases are each addressed in the appropriate order (will always be correct on sample, and should be mentioned).	Positive: This is done well here +3 pts total: clearly states that plan is "logical" or that each task flows from one to next +1.5 pts total: vaguely states that plan is "well organized"
B	Depends on how many problems	The 8 steps are each addressed in the appropriate order (will not be correct on some questions, this should only be mentioned for incorrect aspects).	N/A
C	2	Iteration should be planned into the process.	Negative: No iteration here. Must clearly state that time must be planned in for iterating back to earlier steps when problems are found. (+2 pts)
D	1.5	Relative time allotments should be reasonable: phase II with more time than phase I, phase III leaves enough time not only for building and delays but also for testing (roughly same amount of time as phase II, but depends on project)	Positive: "Getting needs from shop owners" time is reasonable (+0.5 pts) Negative: Too much time spent developing concept before building; "more time for building" (+1 pt)
E	1.5	Gantt chart must have sufficient detail to be useful.	N/A – detail of chart is fine, and comments to that effect should receive 0.5 pts under Step H
F	1.5	Criteria and constraints (i.e., the needs of the project) must be use in analysis, decision-making, and testing.	Positive: Needs are addressed in both concept development and in analysis (1.5 pts)
G	2	Project should be documented throughout (1 pt) with enough time left at end (1 pt) to compile and finish documentation	Negative: Not done well here. +1 pt only if stated that more time is needed for documentation +2 pts total if stated that documentation should occur throughout the process
H	+0.5	Extra credit for insights not listed on rubric	Examples include: "too many things happening in week 4" or "good to be doing more than one thing at a time"
I	-1	Answers that are directly incorrect. (e.g., saying that analysis is not necessary when it is)	

REFERENCES

1. Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. W., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's Taxonomy of educational objectives*. New York, NY: Longman.

2. Bailey, R., Szabo, Z., & Sabers, D. (2004). *Assessing Student Learning about Engineering Design in Project-Based Courses.* Paper presented at the American Society for Engineering Education Annual Conference, Salt Lake City, Utah.

3. Bailey, R., Szabo, Z., & Sabers, D. (2004). *Integrating Education Students in the Assessment of Engineering Courses.* Paper presented at the American Society for Engineering Education Annual Conference, Salt Lake City, Utah.

4. Bloom, S. B., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of educational objectives; the classification of educational goals*. New York, NY: Longmans, Green.

5. Gentile, R. J. (1997). *Educational Psychology* (2nd ed.). Dubuque, IA: Kendall.

6. Nitko, A. J. (2004). *Educational Assessment of Students* (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.

7. Woolfolk, A. (2003). *Educational Psychology* (9th ed.). Boston, MA: Allyn & Bacon.

BIOGRAPHIES

REID BAILEY is an Assistant Professor in the College of Engineering at the University of Arizona. His research interests include engineering design, environmental issues affecting design, and engineering education. He received his B.S. from Duke University and both his M.S. and Ph.D. from the Georgia Institute of Technology.

ZSUZSANNA SZABO is an Adjunct Assistant Professor in the Educational Psychology Department at the University of Arizona. Her research interests include assessment, team learning, and gender issues in education. She received her B. S. in Civil Engineering from Technical University Cluj, B. S. in Psychology from University Babes-Bolyai, both in Romania, her M. Ed. from SUNY at Buffalo, NY, and her Ph.D. from the University of Arizona.