



Creativity in Green Roof Design: Digital Consensual Assessment Outcomes from a High School Engineering Camp

Dr. Jennifer Buelin-Biesecker, North Carolina State University

Dr. Cameron Denson, North Carolina State University

Creativity in Green Roof Design: Digital Consensual Assessment Outcomes from a High School Engineering Camp

Abstract

The field of technology and engineering education has identified creativity as essential to its mission¹. However, a perceived inability to assess creative attributes of students' work has often precluded creativity instruction in the classroom. The Consensual Assessment Technique (CAT) has shown promise in a variety of domains for its potential as a valid and reliable means of creativity assessment. Relying upon an operational definition of creativity and a group of raters experienced in a given domain, the CAT offers the field of engineering education an assessment method that has demonstrated discriminant validity for dimensions of creativity as well as for technical strength and aesthetic appeal. This paper reports on a web-based adaptation of the CAT for rating student projects developed during a week-long engineering camp. High school students were charged with designing a green roof with design constraints related to water management, medium used, plant life, and structural integrity. Images of resulting scale models, technical drawings and poster presentation materials were displayed on a website which was accessed by a team of seven independent raters. Online survey software featuring a series of Likert-type scales was used for ratings. The raters viewed project images on larger computer screens and used iPads to input their assessments. This effort extended the accessibility of the CAT to raters beyond limitations of geographic location.

This paper reports on the inter-rater reliability of the web-based CAT instrument as well as the discriminant validity for purposes of measuring creativity apart from other measured project dimensions. The authors discuss recommendations and implications of this user interface arrangement as it pertains to the practical digital implementation of the CAT.

Introduction

Promotion of creativity and innovation in engineering education is essential to the production of engineers capable of contributing solutions to society's most demanding technological challenges². The need for promoting creative thinking and innovative problem-solving in classrooms has been identified in the research literature³⁻⁴. Moreover, the field of technology and engineering education has identified creativity as essential to its mission¹. Unfortunately, creativity has not always explicitly been part of the goals, objectives and measured results in K-16 classrooms for numerous reasons, including the perceived difficulty in quantifying something that by its very nature is subjective^{5,6}. Lewis⁷ attributed the fledgling state of creative problem-solving assessment to a lack of research on developing ways to help teachers identify and assess inherent creativity in students' design work.

Fostering creativity and creative problem-solving skills can prove challenging amidst classroom expectations of explicit objectives and measurable outcomes. This can be especially difficult within the current goal framework of the average K-12 public school classroom, where engineering education is gaining traction with the release of the Next Generation Science Standards ⁸. Part of the challenge is that teachers may view creative students as “inattentive and disruptive,” tending to “wander away from the regular paths of thought” ⁵ (p. 348). Westby and Dawson ⁶ found significant negative correlations between teachers’ favorite students and their creative students.

Studies have shown, however, that the reliable assessment of creativity in students’ design work is possible ⁹⁻¹¹. This paper proposes the use of the Consensual Assessment Technique (CAT) for creativity assessment in engineering design education.

Engineering Summer Camp

Founded in 1999 as an extension of the Women in Engineering Program, the Engineering Summer Camp offers a week-long engineering camp each summer for 9-12th grade students interested in experiencing engineering, science and technology at a university ¹². Participants in this study attended a multidisciplinary session for rising 9th and 10th grade students. Student campers must pay to participate in the engineering summer camps. Financial aid is provided to those in need. Approximately 90 students were placed in design teams of three students, providing the study with 30 student groups.

Participants in the engineering summer camp were each provided with a camp workbook which contained lead-up activities for the eventual green roof engineering design challenge (Appendix A). At the conclusion of the camp student groups were to provide the staff with a detailed design package to include technical drawings, a working scale model of their design solution, followed by a brief 3-5 minute presentation of their design.

Three high school teachers with backgrounds in science or math were selected as instructors for the engineering summer camp. The instructors were to provide guidance and instruction for the students while facilitating the engineering design experience. Instructors were assisted by undergraduates as well as high school students who served as assistants for the summer camp. The students were divided into groups of three.

The Engineering Summer Camp is housed at a research-one STEM university in the southeast. The activities are developed and coordinated by faculty and staff at the respective university.

Operational Definition of Creativity

When sorting through the profuse definitions and conceptual frameworks available for discussing the concept of creativity, it is useful to identify those most applicable to the task at hand. In this case, the topic of interest is the ability to identify creative attributes of student products. Two types of definitions are useful to this discussion. Hennessey, Amabile, and Mueller¹⁰, whose work in creativity assessment has had tremendous influence upon the design of this study, offered the following:

- *Conceptual definition of creativity*: “A product is considered creative to the extent that it is both a novel and appropriate, useful, correct or valuable response to an open-ended task” (p. 4).
- *Operational definition of creativity*: “A product or response is considered creative to the extent that appropriate observers independently agree that it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated” (p. 4).

Hennessey’s conceptual definition is a useful guide for evaluating student products in technology and engineering education because student products and design processes will vary widely due to many factors and problems are often open-ended. The definition assimilates many prior conceptual definitions¹³ and can be helpful in clarifying to students what is being asked of them when they are told that creativity is a part of their grades. The operational definition establishes the framework and justification for the use of Amabile’s¹⁴ Consensual Assessment Technique (CAT) for evaluating creativity and other dimensions of student responses to open-ended design and problem-solving activities — if knowledgeable raters independently, and with an acceptable level of inter-rater reliability, determine that a student product is creative in its context, then by definition it is. The creative outcomes of student products will be assessed using this method for three major dimensions (creativity, technical strength, and aesthetic appeal) and for nine additional sub-dimensions (novel idea, novel use of materials, complexity, organization, neatness, effort evident, liking, pleasing use of shape/form, and pleasing use of color/value). Factor analysis reveals the CAT’s discriminant validity, in effect revealing whether creativity was measured by raters apart from other characteristics of students’ work.

The Consensual Assessment Technique (CAT)

The CAT is an evaluation tool used by creativity researchers for assessment of creative products by panels of raters. The method is based on the assumption that “a panel of independent raters familiar with the product domain, persons who have not had the opportunity to confer with one another and who have not been trained by the researcher,” are best able to make judgments regarding “the nature of creative products and the conditions that facilitate the creation of those products”¹⁰ (p. 253).

The application of the CAT for making inferences about students' work, and subsequent inferences about pedagogical strategies used in producing that work, depends upon acceptance of an operational definition of creativity: "a product or response is considered creative to the extent that appropriate observers independently agree that it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated" ¹⁰ (p.253). Inter-rater reliability "quantifies the closeness of scores assigned by a pool of raters to the same study participants. The closer the scores, the higher the reliability of the data collection method" ¹⁵ (p. 29).

"In the case of the Consensual Assessment Technique," explained Hennessey, Amabile, & Mueller ¹⁰, "reliability is measured in terms of the degree of agreement among raters as to which products are more creative, or more technically well done, or more aesthetically pleasing than others. [. . .] By definition, interjudge reliability [. . .] is equivalent to construct validity: If appropriate judges independently agree that a given product is highly creative, then it must be accepted as such" (p. 253).

Inter-rater reliability is key to the usefulness of the CAT in classroom evaluation of student work. If stakeholders believe that student work cannot be reliably assessed for creativity because the concept is too enigmatic or inconsistent, then weaving creativity into curricula presents problems for goal setting and measurement. If, however, it can be shown that creativity can be reliably assessed in the classroom, then curricula and education policy can evolve to meet the changing needs of learners. Factors in determining valid and reliable results in the classroom application of the CAT include consideration of the types of raters available to instructors, raters' experience in the given domain, and the number of raters employed.

In early applications of the CAT, Amabile ¹⁴ referred to raters as "experts." What constitutes an appropriate "expert rater" depends upon the researcher's judgment that a rater possesses both knowledge of the domain and "familiarity with the kinds of creative products typically produced by the kinds of subjects in the study" ¹⁶. In recent years researchers have looked at comparisons of novice and expert judgments. At least three categories of raters stand to provide valuable assessment data for engineering design education: self-evaluations conducted by students; peer-evaluations conducted by students enrolled in the same or similar courses; and adult ratings conducted by raters with experience in the domain ^{17, 18}. Across a range of domains, preliminary but significant correlations have been seen between peer evaluations or otherwise non-expert, but somewhat experienced, raters and those made by adult raters with expertise in the domain ¹⁶, suggesting that further investigation could lend insights into greater flexibility of the CAT in classroom practice.

Inter-rater Reliability

Cronbach's alpha is recommended for reporting inter-rater reliability among project raters^{19 & 20, 14, 21-22}. According to Hennessey, Amabile & Mueller¹⁰, who have used Cronbach's alpha for reporting inter-rater reliability results for studies using the CAT, "in most instances, a reliability figure of .70 or higher can be considered evidence of an acceptable level of agreement between judges" (p.256).

The number of judges used can impact the value of the inter-reliability coefficient. The available literature, as well as practical limitations such as time and cost, point to an ideal of approximately seven to 10 raters^{23, 16}.

In order to claim that creativity is being isolated and measured apart from other characteristics of students' work, it is beneficial to demonstrate the instrument's discriminant validity. Items related to creativity will ideally receive consistently different ratings from items related to categorically different types of items. Many studies using the CAT have followed Amabile's¹⁴ three clusters of dimension types: creativity, technical strength and aesthetic appeal, and have included ratings of multiple related sub-dimensions¹⁷. Figure 1 provides a list of sub-dimensions associated with each of the three major dimensions. Factor analysis determines the CAT's discriminant validity; optimally items within each of those three clusters will consistently load together.

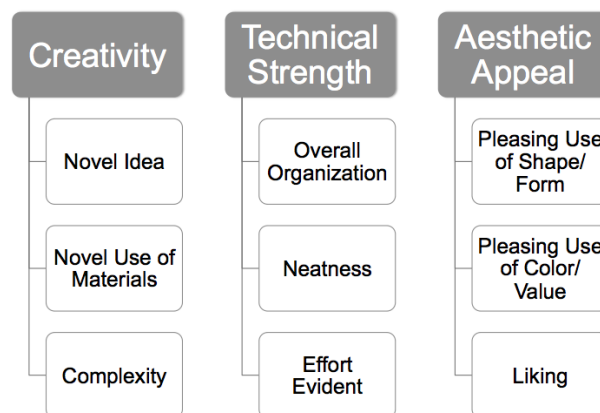


Figure 1. Sub-dimensions associated with each major dimension measured.

Digital CAT interface

Studies conducted using the CAT have traditionally followed similar implementation processes: students create products which are collected by researchers, spread around a single physical space, and viewed and assessed in that space by one rater at a time until the ratings were completed. It may prove valuable to expand the accessibility of consensual assessment beyond the traditional method characterized by displaying student projects throughout a physical space and having raters complete the assessments in person. For this study the researchers developed a web-based assessment interface consisting of 1) an overview video displaying all project images for raters to view prior to the rating session; 2) a website built for the display of project images and documentation; and 3) a web-based version of the consensual assessment instrument, accessed by raters via iPad while viewing the project website on desktop computers.

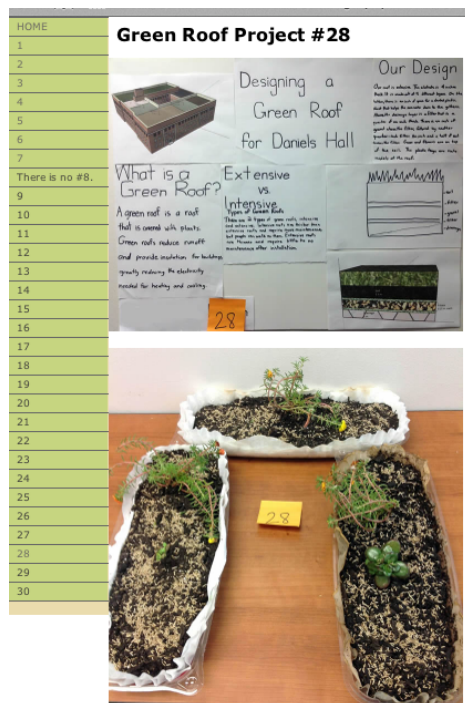


Figure 2. Green roof project website. This figure illustrates the website used by project raters for viewing each of the green roof projects. Photographs of presentation posters and physical models were included for each project.

Figure 3. Consensual assessment instrument for iPad. This figure illustrates the interface used by project raters for making online consensual assessment ratings on twelve dimensions of students' projects.

Methods

The primary research question for this study is whether the digital interface developed for this implementation of the Consensual Assessment Technique would yield strong ($\alpha > 0.75$) inter-rater reliability among the seven raters for the 12 dimensions measured. A secondary question concerning the digital instrument's discriminant validity was also investigated since it is essential to determine whether raters are evaluating creativity apart from other dimensions of projects such as technical strength and aesthetics.

Following an online solicitation for recommendations and volunteers, the authors selected seven raters who were familiar with the engineering design process and experienced in teaching high school-aged students. The raters included: a high school teacher currently teaching Project Lead the Way (PLTW) with over nine years of teaching experience; a professor with joint appointments in engineering and technology education; a National Board certified science teacher with over 19 years teaching experience; a former engineer and current middle school assistant principal; a high school teacher who has taught at the summer engineering camp for five previous years, an engineering camp director with National Board certification as a science teacher, and a 6th grade science teacher with 13 years teaching experience.

Raters were asked to commit approximately two to three hours to a rating session during which they would evaluate student projects on dimensions such as creativity, aesthetic value, and technical strength. Raters were compensated with a \$50 honorarium for their participation.

Procedures

After receiving their team assignments and a brief introduction to the engineering summer camp, student teams received their green roof engineering design challenge on Day 1 of the five-day camp. Each day throughout the week, teams participated in ancillary activities designed to promote critical thinking and problem-solving skills. Student teams were asked to produce a series of modeling artifacts as part of the design requirements. The models that the teams produced included a conceptual model, a mathematical model, a graphical model and a working model illustrating their design solution²⁴. This design process culminated in group presentations to all camp participants, staff and students' families on Day 5.

Following presentations, photographs of students' working models and presentation materials were taken. Images were catalogued by project number on a website built for rater access. Once raters were contracted as participants they were given instructions via email (Appendix B) as well as the project website URL and each rater's unique CAT survey URL.

Results

Distributions of ratings were examined per dimension, and it was determined that no individual's ratings appeared to display systematic bias.

To test inter-rater reliability, Cronbach's alpha was calculated using raters' scores for the 12 separate dimensions rated. It can be seen in Table 1 that all 12 items have reliabilities greater than .70, and that 10 of the 12 have reliabilities greater than .80. This includes creativity, with an inter-rater reliability of 0.86. According to the Landis and Koch²⁵ scale, a reliability coefficient between 0.6 and 0.8 is "substantial" and agreement beyond 0.8 is "almost perfect."

Table 1
Chronbach's α for twelve dimensions measured

Dimensions of Judgment	Cronbach's α
Creativity	0.8642
Aesthetic Appeal	0.8786
Technical Strength	0.7126
Complexity	0.7818
Liking	0.8557
Novel Idea	0.8453
Novel use of Materials	0.8808
Shape/Form	0.8422
Color/Value	0.8914
Organization	0.8269
Neatness	0.8149
Effort Evident	0.8387

In order to evaluate the discriminant validity for this implementation of the CAT, factor analysis was conducted on the mean ratings of the 12 dimensions of judgment (promax rotation). Factor analysis suggested the emergence of three factors, as shown in Table 2, corresponding, albeit not perfectly, with Amabile's¹⁴ paradigm (Figure 1). Although only one factor emerged with an eigenvalue higher than 1.0, consideration of the scree plot (Figure 4) similarly suggests the emergence of three factors, indicated by the rate of change in magnitude of the eigenvalues for factors 1-3. Factor 1 includes creativity and its three subjacent items: novel idea, novel use of materials, and complexity (as well as liking, effort evident, and technical strength). Factor 2 comprises overall aesthetic appeal and its three subjacent dimensions: pleasing use of color/value, pleasing use of shape/form, and liking (as well as novel use of materials and creativity). Factor 3 includes technical strength and two out of three of its subjacent dimensions: overall organization and neatness. This suggests that raters did distinguish between creativity characteristics and other characteristics of the students' green roof designs. It should be noted, however, that factor

analysis is far more stable with larger sample sizes than that of this study, and therefore further testing would be necessary in order to make claims about this instrument's discriminant validity.

Table 2

Factor loading of 12 dimensions, promax rotation

Dimensions of Judgment	Factor 1: Creativity	Factor 2: Aesthetic Appeal	Factor 3: Technical Strength
Creativity	0.61	0.43	0.04
Aesthetic Appeal	0.12	0.79	0.18
Technical Strength	0.79	-0.23	0.41
Color/Value	-0.13	1.02	0.06
Complexity	0.90	0.01	0.08
Effort	0.58	0.21	0.33
Liking	0.55	0.48	0.07
Neatness	0.04	0.18	0.83
Novel Idea	0.88	0.22	-0.09
Novel Materials	0.52	0.57	-0.05
Organization	0.14	0.20	0.72
Shape/Form	0.19	0.75	0.14

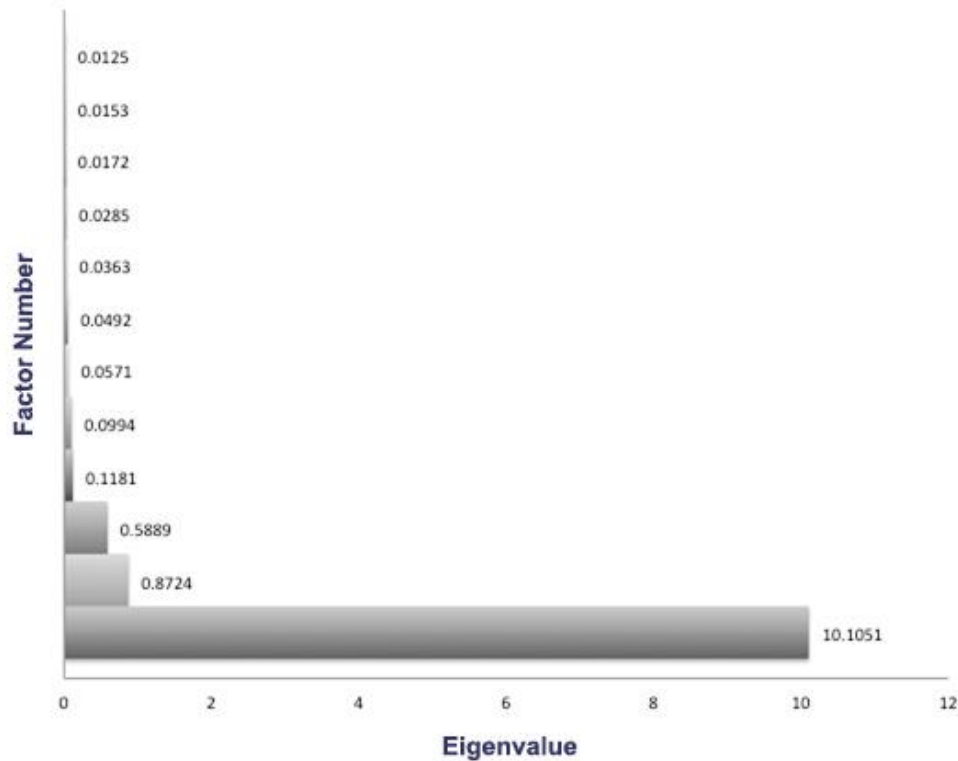


Figure 4. Scree plot for the twelve dimensions measured. This figure illustrates the rate of change in magnitude of the eigenvalues for all twelve components. The slope flattens considerably beyond the third component, suggesting the retention of three factors.

Conclusions

A need for the promotion of creative thinking and innovative problem-solving has been identified in the research literature³⁻⁴, and the importance of creativity in engineering education has become well documented in recent years²⁶. This study builds upon the work of Amabile¹⁴, Hennessey, et.al¹⁰, Hickey¹¹, and others in confirming that creativity can be recognized by raters who are knowledgeable in a domain and that it can be reliably assessed in the classroom. Despite the skepticism various stakeholders (teachers, students, parents, administrators, etc.) have been known to express, a growing body of research supports the assertion that creativity can be reliably recognized and assessed in a formal classroom setting.

The Consensual Assessment Technique shows promise for the assessment of creativity in the domain of engineering design education. Inter-rater reliability among the seven raters was consistently high for all 12 dimensions of judgment measured in this study, and, despite its relative instability with a small sample size, factor analysis suggests that raters were able to recognize and assess creativity apart from other characteristics of student projects. These findings are important to discussions of how curricula and assessment methods might evolve in engineering design education.

The web-based CAT tools used in this study allow instructors to bypass the limitations posed by implementing consensual assessment in a single physical location. The likelihood of obtaining well-qualified raters is improved, and logistical challenges such as displaying a large number of student projects simultaneously are ameliorated.

Further study is needed to develop practical classroom projects and assessment instruments for pre-engineering and engineering students and instructors that will spur students toward meeting their creative potential. One challenge is that the current system can provide raw scores per dimension and project from the slider scale input. The user is required to download and manipulate raw data, and the mean score (between 1 and 9) does not directly translate to a reportable grade. The development of a streamlined software/website template would be beneficial, as this method requires the time, resources, and ability to compile images into an accessible format that is not too cumbersome for raters, and it requires familiarity and access to an online survey instrument.

Larger-scale investigation could be useful in exploring potential benefits of self and peer evaluation to student achievement as well as to classroom creativity assessment. Additional investigation is needed into effective methods for training students to act as peer raters. Consistently high levels of inter-rater reliability found in preliminary cross-domain studies have laid a groundwork for pedagogical investigations comparing, for example, the effects of variables such as design processes, pedagogical strategies, and design prompts on engineering students' creative outcomes. Gender tendencies might also be of interest in similar future studies of larger samples, as prior studies have intermittently shown girls receiving significantly higher creativity scores than boys^{27, 10}.

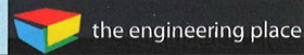
The promotion of engineering students' abilities to think creatively and to effectively communicate their innovative design ideas is fundamentally important. As these findings add to a research base that continues to show creativity can reliably be assessed, engineering instructors are encouraged to include creativity as an explicit objective in their design challenges.

References

1. International Technology Education Association (ITEA/ITEEA). (2000/2002/2007). Standards for technological literacy: Content for the study of technology. Reston, VA.
2. Felder, R.M. (1987). On creating creative engineers. *Engineering Education* 77(4), 22-227.
3. Todd, S. M., & Shinzato, S. (1999). Thinking for the Future: Developing Higher-Level Thinking and Creativity for Students in Japan—and Elsewhere. *Childhood Education*, 75(6), 342-345.
4. National Center on Education and the Economy (2010). *Equipping the Federal Government to Counter Terrorism*. Retrieved September 7, 2011 from <http://www.nap.edu/catalog/10415.html>.

5. Lau, S., & Li, W. (1996). Peer status and perceived creativity: Are popular children viewed by peers and teachers as creative? *Creativity Research Journal*, 9(4), 347-352.
6. Westby, E. L., & Dawson, V. L. (1995). Creativity: Asset or burden in the classroom?. *Creativity Research Journal*, 8(1), 1-10.
7. Lewis, T. (2005). Creativity: A framework for the design/problem solving discourse in technology education. *Journal of Technology Education*, 17(1), 36.
8. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS. (2013). Next Generation Science Standards.
9. Amabile, T.M. (1996). *Creativity in context*. Boulder, CO: Westview Press.
10. Hennessey, B.A., Amabile, T.M., & Mueller, J.S. (2011). Consensual assessment. In: Runco, M.A. & Pritzker, S.R. (Eds.) *Encyclopedia of Creativity*, Second Edition, Vol. 1, pp. 253-260. San Diego: Academic Press.
11. Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, 49(3), 234.
12. The Engineering Place. Retrieved from <http://www.engr.ncsu.edu/theengineeringplace/summerprograms/high-school/>. November 14, 2013.
13. Cropley, A.J. (1999). Definitions of creativity. In M. Runco and S. Pritzker (Eds.) *Encyclopedia of creativity: Vol. 1*. Academic Press.
14. Amabile, T.M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.
15. Gwet, K. L. (2008). Intrarater reliability. *Wiley encyclopedia of clinical trials*.
16. Kaufman, J.C., Baer, J., Cole, J., & Sexton, J.D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171-178.
17. Buelin-Biesecker, J.K., & Wiebe, E.N. (2013). Can pedagogical strategies affect students' creativity? Testing a choice-based approach to design and problem-solving in technology, design, and engineering education. *Proceedings of the 2013 American Society for Engineering Education Annual Conference & Exposition*, Paper ID #6045.
18. Kaufman, J.C., Russell, C.M., & Plucker, J.A. (2013). Models and methods of assessing creativity. In Saklofske, D. H., Reynolds, C. R., & Schwan, V. L. *The Oxford handbook of child psychological assessment*. Oxford: Oxford University Press.
19. Uebersax, J. (2010). *Kappa coefficients*. <http://www.john-uebersax.com/stat/kappa.htm>.
20. Uebersax, J. (2010b). *The tetrachoric and polychoric correlation coefficients*. <http://www.john-uebersax.com/stat/tetra.htm>.
21. Gwet, K. (2010). *Handbook of Inter-rater Reliability*. Gaithersburg, MD: Advanced Analytics.
22. Gwet, K. (2011). *On the Krippendorff's alpha coefficient*. Manuscript submitted for publication. Retrieved October 2, 2011 from http://agreestat.com/research_papers/onkrippendorffalpha.pdf.
23. Fiske, H.E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education*, 25, 256-263.
24. Lammi, M. D., & Denson, C. (2013). Pre-service Teachers' Modeling as a Way of Thinking in Engineering Design. *Proceedings of the 2013 American Society for Engineering Education Annual Conference & Exposition*, Paper ID #5867.
25. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
26. Amato-Henderson, S., Kemppainen, A., & Hein, G. (2011). Assessing creativity in engineering students. 41st ASEE/IEEE Frontiers in Education Conference. Retrieved from <http://fie-conference.org/fie2011/papers/1440.pdf>.
27. Amabile, T.M. (1983). *The social psychology of creativity* (Vol. 11). New York: Springer-Verlag.

Green Roof Design Challenge



Objective: To design an intensive green roof for Daniels Hall, that will absorb rainwater and provide insulation for the building. You can assume that the roof has been reinforced to hold 100 lbs per square foot. In addition this green roof should provide a beautiful, natural green place that students, faculty, and visitors can enjoy.

While you will not actually build and plant this full size green roof in the next five days, you will do the following:

1. Create a very detailed design, complete with technical drawings, that could be used to create an actual green roof on top of Daniels Hall.
2. Create a working scale model of your final design.
3. Prepare a brief 3 – 5 minute presentation about your design to be given to your parents on Friday afternoon.

Materials for scale model:

- Soil
- Sand
- Pebbles, Gravel, and Rocks
- Various seeds
- Various plants
- Various filter materials
- Various plastic materials
- Plastic tubing
- Scrap wood
- Duck tape
- Various types of glue
- Various recycled items
- Any other reasonable requests



Tools:

- Scissors, Saws, & Drills
- Thermometer
- Scales

How will you accomplish all of this in one week?

Hard work and lots of fun. We'll help you out by giving you daily goals. See the next page.

Appendix B

Instructions to Raters

Dear [Rater],

Thank you once again for agreeing to serve as a rater for this study. This email contains the instructions and links you will need to complete the ratings. Please read the entire email before you begin rating projects.

*Data collection for this study is entirely web-based. Project information and photographs are located at [REDACTED]. The home page contains a video and a design brief. On the left side of the screen you will see links to all student projects, numbered 1 to 30 (*there is no #8). Each project page contains images of the students' posters and models.*

Please begin the rating process by reading the design brief and viewing the short video on the home page. (I recommend using Safari to view the website; if you have any trouble viewing the video you may also access it here:

[REDACTED]) This video is an overview of the images you will find on the website. It serves as an introduction to the products created by the students, and it will give you a sense of the range of abilities represented in the sample. It is essential to our methodology that you look over all the products prior to rating any projects, and that you rate projects relative to each other rather than making ratings based on some absolute standard. In other words, consider what the camp students were able to do given time, instruction, supplies, etc., rather than what you think they should be able to do.

Ratings will be made on an iPad using Qualtrics online survey software while you view the project images on your computer screen. Please let me know if you would like for me to lend you an iPad or if you would like to come to [campus] to use a work station here. Here is the link to your survey: [REDACTED]. This URL is unique to you (Rater #3).

Each project will be rated on 12 dimensions ("technical strength," "novel idea," "creativity," etc.). A brief description of the dimension is given, but beyond that description we are relying on your expert judgment rather than further definitions or rubrics.

Simply open the survey using the link provided above, enter the project number you are currently viewing (1-30), move the sliders to indicate your rating for each of the 12 dimensions, and submit using the button at the bottom of the survey.

Once you submit for the first project, a "thank you" message will appear, and beneath it will appear the survey link. By clicking the link, you will be taken to a fresh, reordered survey form. This is how you will rate the next project. Please repeat this process for all 29 projects.

**It may be helpful to write out the numbers 1 through 30 and to mark off each number when you submit ratings for that project. Once you're finished with your ratings, I'll verify that you didn't skip any projects. I'll simply email you if you missed one, and you can return to the website to rate any you might have missed.*

Please feel free to contact me via email or phone [REDACTED] if you have any questions or experience any technical difficulties.

We would appreciate your having ratings completed by October 1 if possible. Please let me know if that will not fit your schedule. We understand that you are very busy. After the ratings are complete we will ask you to complete a brief demographics survey which will also be used to collect the information we need in order to process your \$50 compensation as quickly as possible. We will also be happy to share the results of this study with you.

Thank you for your contribution to this work!