Consensual Assessment: A Means of Creativity Evaluation for Engineering Graphics Education

Jennifer K. Buelin-Biesecker
Department of STEM Education
North Carolina State University

Cameron D. Denson
Department of STEM Education
North Carolina State University

Abstract

The perceived inability to assess creative attributes of students' work has often precluded creativity instruction in the classroom. This digest discusses the use of the Consensual Assessment Technique (CAT) for assessment of creative products, emphasizing its suitability to engineering graphics education. A brief methodology of the instrument is offered including techniques for ensuring the validity and reliability of its results. A practical example of using the CAT to assess engineering graphics is provided. The digest concludes with a call for further studies expanding upon the growing body of research that addresses creativity assessment in the classroom.

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 challenges (Felder, 1987). In addition to cognitive operations involved in engineering design processes-- not least among them the ability to "accurately perceive the visual-spatial world and transform these perceptions" into graphical plans (Harris & Meyers, 2007)-- engineering students must learn to effectively "communicate their designs and engineering analyses using both verbal and graphical languages" (Ault, 2002). Hansberry and Lopez (2005) identified "graphic illiteracy" as a source of communication breakdown in the engineering field. Central to the development of graphic literacy, they said, lies "the ability to combine creativity and visualization to make unique designs."

Fostering creativity and creative problem-solving skills can prove challenging amidst the 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, a context in which engineering education is gaining traction with the release of the Next Generation Science Standards (Achieve, Inc., 2013). 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" (Lau

& Li, 1996, p. 348). Westby and Dawson (1995) found significant negative correlations between teachers' favorite students and their creative students.

The need for promoting creative thinking and innovative problem-solving in classrooms has been identified in the research literature (Todd & Shinzato, 1999; National Center on Education and the Economy, 2010). Moreover, the field of technology and engineering education has identified creativity as essential to its mission (ITEA/ITEEA, 2000/2002/2007). 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 assessing it (Lau & Li, 1996; Westby & Dawson, 1995). Lewis (2005) 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.

Studies have shown, however, that the reliable assessment of creativity in students' design work *is* possible (Amabile, 1996; Hennessey, Amabile, & Mueller, 2011; Hickey, 2001). This paper proposes the use of the Consensual Assessment Technique (CAT) for creativity assessment in engineering design graphics education.

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" (Hennessey, Amabile, & Mueller, 2011, 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" (Hennessey, Amabile, & Mueller, 2011, 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" (Gwet, 2008, p. 29). "In the case of the Consensual Assessment Technique," explained Hennessey, Amabile, & Mueller (2011), "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 making claims about 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 (1983) 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" (Kaufman, Baer, Cole, & Sexton, 2008). 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 technical graphics 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 (Buelin-Biesecker & Wiebe, 2013; Kaufman, Russell, & Plucker, 2013). Across a range of domains, preliminary but significant correlations have been seen between peer evaluations or otherwise nonexpert, but somewhat experienced, raters and those made by adult raters with expertise in the domain (Kaufman, Baer, Cole, & Sexton, 2008), suggesting that further investigation could lend insights into greater flexibility of the CAT in classroom practice.

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 ten raters (Fiske, 1977; Kaufman, Baer, Cole, & Sexton, 2008).

Inter-rater Reliability and Discriminant Validity

Agreement among raters has been reported using several different coefficients for inter-rater reliability. Debate as to the best coefficient for tests using multiple raters is abundant in the literature and is beyond the scope of this digest. Cronbach's α is recommended for reporting inter-rater reliability (Uebersax, 2010a & 2010b; Amabile, 1983; Gwet, 2010 & 2011). According to Hennessey, Amabile, & Mueller (2011), who have used Cronbach's α, "in most instances, a reliability figure of .70 or higher can be considered evidence of an acceptable level of agreement between judges" (p.256).

In order to claim that creativity is being isolated and measured apart from other characteristics of students' work, it is essential to demonstrate an 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's (1983) three clusters of dimension types: creativity, technical strength and aesthetic appeal, and have included ratings of multiple related subdimensions (Buelin-Biesecker & Wiebe, 2013). Factor analysis determines the CAT's discriminant validity; optimally items within each of those three clusters will consistently load together.

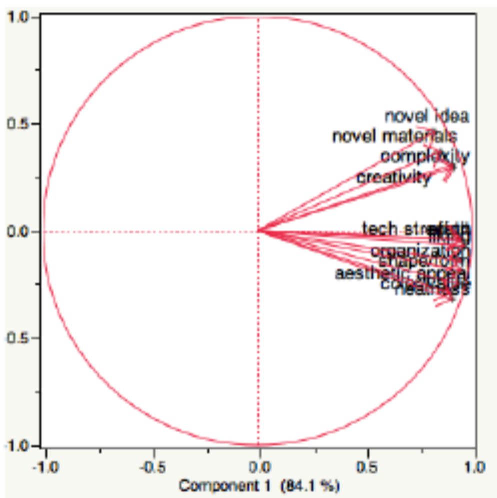

Figure 1. Factor analysis reveals the emergence of a 'creativity cluster' comprising creativity and the three related subdimensions of creativity: novel idea, novel use of materials, and complexity. This factor loading indicates discriminant validity for creativity; i.e., raters judged creativity and related subdimensions apart from dimensions of technical strength and aesthetic appeal (Buelin-Biesecker & Wiebe, 2013).

Consensual Assessment in Practice: Assessment in the Technical Graphics Classroom

The CAT is adaptable to a range of formative and summative assessment situations in technical graphics education. Engineering students' undergraduate matriculation usually culminates with a capstone engineering design experience. Given a design challenge, students develop a solution that meets prescribed design criteria and stays within design constraints. It is imperative that students effectively use graphical representations to communicate their designs in order to explicate the functionality and features of their solution. Many instructors are adept at assessing technical aspects of student designs but the ambiguity often associated with creativity presents teachers with an assessment conundrum. The CAT provides a valid, reliable tool to assess creative aspects of design solutions.

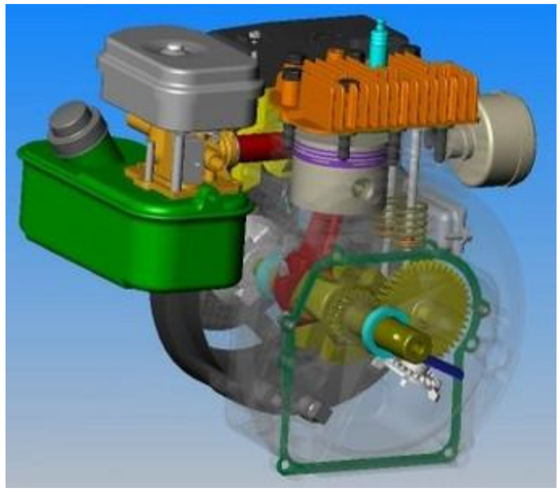

Figure 2. A final project from a capstone engineering design graphics course is a suitable product for evaluation using the Consensual Assessment Technique. Major dimensions measured include creativity, technical strength, and aesthetic appeal. (Courtesy of Prof. T. Branoff, North Carolina State University.)

Limitations of the CAT

Classroom adaptation of the CAT can be limited by time constraints, expense and logistical challenges. Suitable projects must be open-ended enough to allow creative solutions. Appropriate raters must be procured and possibly compensated. Raters must work independently and must have access to all products at once. Statistical analysis is a necessary step in interpreting results. However, these challenges can be overcome with thoughtful planning, particularly in higher education.

Conclusions and Recommendations

A growing body of work supports the assertion that creativity can be reliably assessed using the CAT and that the method is appropriate for the domain of technical graphics education. The need to promote engineering students' abilities to think creatively and to effectively communicate their innovative ideas graphically is fundamentally important.

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.

References:

Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS. (2013). Next Generation Science Standards.

Amabile, T.M. (1983). *The social psychology of creativity.* New York: Springer-Verlag.

Amabile, T.M. (1996). *Creativity in context.* Boulder, CO: Westview Press.

Ault, H.K. (2002). Engineering design graphics as a communications tool for mechanical design: A broader view. *Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition*, Session 2238.

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.

Felder, R.M. (1987). On creating creative engineers. *Engineering Education, 77*(4), 22-227.

Fiske, H.E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education, 25*, 256-263.

Gwet, K. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical & Statistical Psychology*, *61*(1), 29-48.

Gwet, K. (2010). *Handbook of Inter-rater Reliability.* Gaithersburd, MD: Advanced Analytics.

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.

Hansberry, E.W., & Lopez, G.W. (2005). Avoiding graphic illiteracy: Incorporating architectural graphics into the engineering curriculum. *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition*.

Harris, L.A., & Meyers, F. (2007). Graphics: Into the 21st century. *Proceedings of the 2007 American Society for Engineering Education Annual Conference & Exposition*, Session 327.

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.

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.

International Technology Education Association (ITEA/ITEEA). (2000/2002/2007). *Standards for Technological Literacy: Content for the Study of Technology*. Reston, VA.

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.

Kaufman, J.C., Russell, C.M., & Plucker, J.A. (2013). Models and methods of assessing creativity. In Saklofske, D. H., Reynolds, C. R., & Schwean, V. L. *The Oxford handbook of child psychological assessment*. Oxford: Oxford University Press.

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.

Lewis, T. (2005). Creativity—A framework for the design/problem solving discourse in technology education. *Journal of Technology Education, 17*(1), 36-53.

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.

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), 342-45.

Uebersax, J. (2010). *Kappa coefficients.* http://www.john-uebersax.com/stat/kappa.htm.

Uebersax, J. (2010b). *The tetrachoric and polychoric correlation coefficients*. http://www.john-uebersax.com/stat/tetra.htm.

Westby, E.L., & Dawson, V.L. (1995). Creativity: Asset or burden in the classroom? *Creativity Research Journal*, *8*(1), 1-10.