Reviewing Approaches to Assessment of Students' Designs

Douglas B. Clark*, Werklund School of Education, University of Calgary, douglas.clark@ucalgary.ca
Chris Ostrowdun*, Werklund School of Education, University of Calgary, chris.ostrowdun@ucalgary.ca
Stefan Rothschuh*, Werklund School of Education, University of Calgary, stefan.rothschuh@ucalgary.ca
Senay Purzer, School of Engineering Education, Purdue University, senay@purdue.edu
*The first three authors are listed alphabetically and are equal coauthors

Abstract: The structure of assessments can affect what is valued, taught, and learned by teachers and students. It is therefore of great importance to review the approaches proposed for assessing student design (processes and outcomes) given the increasing focus on design in STEM curricula. We began with the goal of reviewing such approaches across 17 prominent journals that address STEM and design education. Database searches resulted in 2101 raw hits and ultimately resulted in a sample of 27 articles, predominantly in engineering education. We performed a content analysis to identify key assessment practices. The majority (20) included a focus on performance of the design, and a substantial number included a focus on communication (15) and scoping (11). While less prevalent, divergent thinking (9), creativity (9), convergent thinking (8), and collaboration (8) were also broadly represented. Ethical considerations were not strongly represented in the reviewed assessments, although there were important exceptions.

Introduction

In STEM education, there is increased attention to "design," both as a disciplinary practice and a pedagogical approach. Given the interdisciplinary nature of design, research on design learning and assessment has been conducted in an array of disciplinary journals. Our goal was to identify peer-reviewed articles about design assessment and to analyze how educators assess students' designs in STEM and design education. We explored the following questions: (a) what disciplines are involved, (b) what aspects of the design process and/or outcomes are assessed, (c) what approaches are used to assess design, (d) at what granularity are the assessments conducted, (e) who is involved in the assessment process, and (f) what age groups of students are the focus. To evaluate the status of assessment of design in STEM education, we performed a systematized literature review (Grant & Booth, 2009) and content analysis across seventeen leading journals that address STEM and design education based on Journal Citation Reports (JCR), Scientific Journal Rankings (SJR), and h-index statistics.

Theoretical framing

The structure, content, and scope of assessment drive what is valued, taught, and learned by teachers and students (e.g., Entwhistle & Peterson, 2004; Jürges, Schneider, Senkbeil, & Carstensen, 2012). It is therefore of great importance to review the approaches proposed for assessing design given its multifaceted nature and the increasing role of design in STEM curricula. Design educators, recognizing the importance and the challenges in assessing informed design practices, have developed instruments for assessing students' design processes (Atman, Chimka, Bursic, & Nachtmann, 1999; Bailey & Szabo, 2007). In design education, design outcomes (products, artifacts, or systems) also play an important role. Design outcomes need to achieve high quality while attending to competing criteria and constraints. Purzer, Fila, and Nataraja (2016) introduce the concept of trade-off value when evaluating student design artifacts. One of the challenges student designers face is balancing competing design criteria such as feasibility and novelty (Oman, Tumer, Wood, & Seepersad, 2012; Shah, Woodward, & Smith, 2013). Studies show that undergraduate engineering students skew toward novelty in their first year in college, while the balance switches toward technical feasibility in the final years (Genco, Hölttä-Otto, & Seepersad, 2012). The purpose of this review is to better understand how assessment practices attend to these multifaceted aspects of design processes and design outcomes.

Methods and search criteria

We bounded our review to articles published between 2002 through August 2018. Given the common use of the terms "design" and "assessment" in research beyond the scope of our study, we identified 17 specific journals as opposed to a broad search within selected databases. We searched across four databases (i.e., ASC, ERC, ERIC, and AA) indexed the 17 specific journals we targeted. We used the following search terms and Boolean operators at the abstract and title level: (Design* OR Ideation OR Creativity) AND (Assess* OR Rubric OR Framework OR (Coding Scheme) OR Measure OR Metric OR Instrument OR Method). This scan resulted in 2101 raw hits

across the 17 journals. Our team identified several inclusion criteria in iterative discussions, including the proposition, presentation, or validation of an assessment framework, or sufficient theoretical warranting (a full list of criteria will be included in the journal publication). Articles that straddled the borders of the inclusion criteria were reviewed at the abstract level by the group for a final decision. We then reviewed each article and evaluated whether the assessment instruments therein were described in sufficient detail to meet the criteria. The literature search resulted in a sample of 27 articles: 23 from engineering education journals and 4 from a design journal. We performed a content analysis of the final 27 articles to identify key components of assessment.

Results and interpretations

In our inductive coding, categories arose around multiple aspects of design, including: (a) problem scoping, (b) divergent thinking, (c) creativity, (d) design performance and functionality, (e) convergent thinking, (f) collaboration, and (g) communication. While these seven categories were generally distinct, the clearest distinctions involved design performance, communication, and collaboration, while the distinction between divergent thinking and creativity was the least clear in many cases. This is complicated by terminology that is often used interchangeably in the literature (e.g., originality and novelty).

Table 1: Coding Results Overall And By Article

	Totals		Acar (2004)	Azmy & Mokhtar (2017)	Bar-Eli (2013)	Bartholomew et al. (2018)	Berry & Carlson (2010)	Charyton & Merrill (2009)	Charyton et al. (2011)	Chiaradia, et al. (2017)	Dancz et al. (2017)	Davis et al. (2002)	Davis et al. (2010)	Demirkan & Afacan (2012)	Guzzomi et al. (2015)	Jaeger & Adair (2015)	Lans & Verkroost (2004)	Mccormack et al. (2015)	Oehlers (2006)	Ozaltin et al.(2015)	Reid & Cooney (2008)	Ringwood et al. (2005)	Sherrett et al. (2013)	Sluis-Thiescheffer et al. (2016)	Steiner et al. (2011)	Sung-Hee et al. (2015)	Thompson et al. (2013)	Watson et al. (2017)	Welch et al. (2009)
Design Focus																													
Scoping	11			1				1	1			1					1	1		1	1		1		1	1			
Divergent Think	9				1			1	1			1				1	1			1				1		1			
Creativity	9							1	1			1		1					1	1		1		1		1			
Convergent Think	8			1				1	1			1					1	1					1			1			
Performance	20			1	1	1		1	1	1	1	1			1	1	1	1	1	1	1	1	1		1	1		1	
Collaboration	8											1	1			1		1			1				1	1			1
Communication	15		1	1			1				1	1	1			1	1	1		1	1				1	1	1		1
Total Design Foci			*1*	*4*	*2*	*1*	*1*	*5*	*5*	*1*	*2*	*7*	*2*	*1*	*1*	*4*	*5*	*5*	*2*	*5*	*4*	*2*	*3*	*2*	*4*	*7*	*1*	*1*	*2*
Students																													
K12	1																							1					
Undergraduate	25		1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1
Graduate	1									1																			
Evaluators																													
Instructor/TA	25		1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1
Peer	12		1	1			1			1			1		1	1		1			1	1			1	1			
Self	9		1				1					1	1			1		1			1				1	1			
External	9		1			1		1	1	1				1				1		1								1	
Total Evaluators			*4*	*2*	*1*	*2*	*2*	*2*	*2*	*3*	*1*	*2*	*3*	*2*	*2*	*3*	*1*	*4*	*1*	*2*	*3*	*2*	*1*	*1*	*3*	*3*	*1*	*1*	*1*

Most of the assessment approaches addressed multiple categories (2.96 out of 7 categories on average). Among the 27 articles, the majority (20) included a focus on assessing design performance, and a substantial number included a focus on assessing communication (15) followed by problem scoping (11). While less prevalent, divergent thinking (9), creativity (9), convergent thinking (8), and collaboration (8) were also broadly represented (Table 3). Notably, most articles assessed the performance or function of a final design outcome but the processes of design were not always assessed. We speculate the logistics of assessing a final design are often more manageable for instructors than assessing intangible or real-time aspects of the design process in the midst of running a class or perhaps the assessment of design processes is done in implicit ways with a formative intent. Undoubtedly, the processes of design are as important as final projects as argued in prior research (see Atman et

al.), but we recognize that the pragmatic realities of facilitating and teaching a class may challenge the bandwidth to conduct simultaneous real-time assessments of process. There is work underway to collect real-time process data digitally to automate assessment of process (e.g., Worsley & Blikstein, 2016; Xie et al., 2014), which may change the nature of assessment of design radically in the future, but these approaches are not yet scalable. These digital tools can help complement, extend, or add depth to what is currently feasible.

One major influence on assessment foci are professional engineering organizations, particularly the Accreditation Board for Engineering and Technology (ABET). Within the 2019-20 ABET definition of Engineering Design, we see a strong alignment with the aspects of design highlighted by our coding of the assessments in terms of problem scoping (e.g., identifying opportunities and developing requirements), creativity (e.g., creative decision-making processes), design performance and functionality (e.g., meeting needs and specifications), ideation (e.g., generating multiple solutions), convergent thinking (e.g., evaluating solutions against requirements). Collaboration (e.g., ability to function effectively on a team) and communication (e.g., ability to communicate ideas and technical information effectively with a range of audiences) are key ideas represented in the specified ABET criteria for student outcomes. Interestingly, creativity is not specified within the 2019-20 student outcome criteria, although it is included within the definition of engineering design. That said, historically creativity has been even less represented in the criteria and definitions.

Aspects of ABET that are less extensively represented in the reviewed assessments include conducting experiments, applying science and mathematics, and information literacy/life-long learning. Engineering ethics and ethical considerations seem to be the very least represented, although there are exceptions. Berry (2010), for example, focused heavily in their approach to Calibrated Peer Review on having the students write a social impact document using the IEEE Code of Ethics as the rubric. Other assessments may include some aspect of attending to ethical issues as part of the task but not as a major aspect of the assessment rubrics themselves. While beyond the scope of this review, we find such omissions troubling given the very real impacts design can have on people's lives. For example, designers can face heavy ideological and ethical dilemmas when engineering products that are used in warfare (Philip, Gupta, Elby, & Turpen, 2018) or to marginalize specific populations (Chellew, 2019). This seems to be an area for future growth in the assessments. While not an assessment of students' actual designs, the approach outlined by Christensen, Hjorth, Iversen, and Blikstein (2016) assesses students' stance toward inquiry in a way that attends to the complexity inherent in wicked problems in terms of social and ethical dimensions among others.

In terms of how assessment is conceptualized and operationalized, most articles described summative rather than formative assessment procedures. We surmise this stems from a combination of formative assessments often being less formal (e.g., undocumented verbal commentary from an instructor circulating among students), traditional preferences to summative assessments in STEM education, and an interest in supporting accreditation processes (e.g., ABET). For demographics, we anticipated undergraduate education would be the predominant focus of the assessments but were surprised by the limited focus at the K-12 level given the increasing interest and emphasis in K-12 education on design, design thinking, and engineering. There are many excellent studies of K-12 student design that are not focused on "assessment" or where design-based learning uses design as a pedagogy and hence their assessment focuses on student learning of science concepts (e.g., Crismond, 2001; Hmelo, Holton, Kolodner, 2000; Kolodner et al., 2003.). In terms of evaluators, there was the integration of multiple types of assessor (2.04 on average). Instructors and TAs were involved in almost all of the assessments (25 of 27), but most (18) of the assessments also involved other people including peers (12), self-assessments (9), and external disciplinary experts (9).

In terms of approach, the assessments primarily depend on the application of rubrics. Some articles do not provide enough information about their rubrics to be reliably applied by others, but many of the articles engage in a high degree of granularity with highly descriptive rubrics. Some of the assessments leverage complex approaches and mathematics within the assessment to create a final numerical score with great precision, but this apparent precision is necessarily limited because the input numbers are based, at times, on highly subjective and underspecified criteria. Another group of the assessments are more subjective in their orientation from beginning to end. Sometimes this has the advantage of not artificially systematizing the process, but greater specification within the rubric criteria would likely most improve the rigor and reliability of the assessment processes we reviewed. In addition to the scaled rubric approaches, some assessments pursued alternative mechanics, such as establishing win-loss ratios over the course of multiple paired comparisons. Motivations for alternative approaches included increased efficiency in grading, greater depth of learning, greater engagement of students, and greater authenticity with professional practice.

Related to this, we noticed a considerable range in the granularity of assessments. For instance, Charyton and Merill (2009) used single-words to describe originality criteria, while Davis et al., (2002) used detailed criteria and sub-criteria in their assessment. We find benefits and drawbacks to both approaches. The simplified or open-

ended descriptors afford flexibility and space for amorphic aspects of design such as idea generation and creativity, but they also limit consistency in how to carry out assessment procedures and how students understand what is expected of them. In contrast, detailed criteria reduce ambiguity in what is being assessed and how but also make design more rigid, which can run against more freeform or fluid notions of innovation and novelty.

Educational importance of the study

In our grounded coding, most of the assessment approaches focused on multiple areas. Among the 27 articles, the majority included a focus on performance assessment, and a substantial number included a focus on communication and problem scoping. While not as prevalent, divergent thinking, creativity, convergent thinking, and collaboration were also broadly represented. Engineering ethics and ethical considerations seem to be the very least represented, although there are exceptions. What is assessed greatly influences what is taught and learned, and careful attention to these less represented aspects of design will therefore be an important area for future development for assessment as design takes on an increasingly central role in STEM education.

Selected references

Acar, B. S. (2004). Analysis of an assessment method for problem-based learning. *European Journal of Engineering Education, 29*(2), 231–240. https://doi.org/10.1080/0304379032000157213

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design studies, 20*(2), 131-152.

Azmy, A., & Mokhtar, H. (2017). A Project-Based Framework for Teaching and Assessment of Design Modules. *International Journal of Engineering Education, 33*(2), 629-638.

Bailey, R., & Szabo, Z. (2007). Assessing engineering design process knowledge. International Journal of Engineering Education, 22(3), 508.

Berry, F. C., & Carlson, P. A. (2010). Assessing engineering design experiences using calibrated peer review. *International Journal of Engineering Education, 26*(6), 1503-1507.

Charyton, C., & Merrill, J. A. (2009). Assessing general creativity and creative engineering design in first year engineering students. *Journal of Engineering Education, 98*(2), 145-156. https://doi.org/10.1002/j.2168-9830.2009.tb01013.x

Chellew, C. (2019). Defending Suburbia. *Canadian Journal of Urban Research*, *28*(1), 19-33.

Christensen, K. S., Hjorth, M., Iversen, O. S., & Blikstein, P. (2016). Towards a formal assessment of design literacy: Analyzing K-12 students' stance towards inquiry. *Design Studies*, *46*, 125-151.

Crismond, D. (2001). Learning and using science ideas when doing investigate-and-redesign tasks: A study of naive, novice, and expert designers doing constrained and scaffolded design work. *Journal of Research in Science Teaching*, *38*(7), 791-820.

Davis, D. C., Gentili, K. L., Trevisan, M. S., & Calkins, D. E. (2002). Engineering Design Assessment Processes and Scoring Scales for Program Improvement and Accountability. *Journal of Engineering Education*, *91*(2), 211-221.

Entwistle, N., & Peterson, E. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments. *International Journal of Educational Resesearch, 41*(6), 407–428. https://doi.org/10.1016/j.ijer.2005.08.009

Genco, N., Hölttä-Otto, K., & Seepersad, C. C. (2012). An experimental investigation of the innovation capabilities of undergraduate engineering students. *Journal of Engineering Education*, *101*(1), 60-81.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hmelo, C.E., Holton, D.L., Kolodner, J.L. (2000). Designing to Learn about Complex Systems. Journal of the Learning Sciences, Vol. 9, No. 3, pp. 247 - 298.

Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, *31*(1), 56-65. https://doi.org/10.1016/j.econedurev.2011.08.007

Philip, T. M., Gupta, A., Elby, A., & Turpen, C. (2018). Why ideology matters for learning: A case of ideological convergence in an engineering ethics classroom discussion on drone warfare. *Journal of the Learning Sciences*, 1–41.

Purzer, S., Fila, N., & Nataraja, K. (2016). Evaluation of current assessment methods in engineering entrepreneurship education, *Advances in Engineering Education, 5*(1), 1-27.

Worsley, M., & Blikstein, P. (2016). Reasoning Strategies in the Context of Engineering Design with Everyday Materials. *Journal of Pre-College Engineering Education Research, 6(*2), 57–74.