Development of a scenario-based instrument to assess co-design expertise in humanitarian engineering

Scott Daniel and Andrea Mazzurco

Engineering Practice Academy, Swinburne University of Technology, Melbourne, Australia

sdaniel@swin.edu.au

Development of a scenario-based instrument to assess co-design expertise in humanitarian engineering

Co-design is fundamental to humanitarian engineering and increasingly recognised as such in engineering curricula. However, it is challenging to teach, learn, and assess. In this paper we describe the development and validation of a scenario-based instrument to distinguish novice and expert approaches to co-design in the context of humanitarian engineering. The instrument assesses the extent to which respondents describe stakeholder participation in each of the *scope*, *design*, and *deliver* phases of the design process, with co-design experts taking a collaborative approach throughout. We analyse and compare responses to the instrument from first-year undergraduate engineering students and experienced humanitarian engineering practitioners. Implications for educators, to use this scenario-based assessment in their own research, teaching, and curriculum development, are discussed in detail.

Keywords: co-design; scenario-based assessment; humanitarian engineering; rubric; novice-expert comparisons

Introduction

Humanitarian engineering (HE) is an emerging engineering field that has been defined as "design under constraints to directly improve the wellbeing of underserved populations" (Mitcham and Muñoz (2009), p. 191). As such, it entails the design of appropriate technologies and services in multiple sectors, including agriculture, energy, habitat, health, water and many others (Hazeltine and Bull 1999). In some contexts, it also refers to engineering work for disaster relief (Turner, Brown et al. 2015). In recent years, HE has become more popular and institutions across the world have started offering majors and minors in humanitarian engineering or similar fields (e.g., Bratton (2014), Moskal, Skokan et al. (2008), Passino (2009), Smith, Mazzurco et al. (2018)). As interest in such programs increases, it becomes imperative to understand what skills, knowledge, and attributes graduates need to possess.

Insight into what capabilities are needed to perform effectively in HE projects can be drawn from studies looking at principles and best practice in HE. Based on a review of over 200 publications, Mattson and Wood (2014) identified nine main principles, which underline the importance of involving local communities in the design process, adapting technologies and project management approaches to the specific socio-cultural and environmental context, being aware of gender dynamics, and more. The importance of context and community involvement also emerges from analysing case-studies of failed projects (Mazzurco and Jesiek 2014, Wood and Mattson 2016) and has been echoed by multiple scholars (Lucena and Schneider 2008, Schneider, Leydens et al. 2008, Mazzurco and Jesiek 2017). Thus, from these studies, it is evident that co-design expertise, or the ability to involve the local community across all stages of the design process, is one of the most important capabilities for humanitarian engineers.

However, although there has been extensive research investigating key facets of design expertise and the development of design skills in traditional engineering fields (e.g., Atman, Adams et al. (2007), Kilgore, Atman et al. (2007), McKenna (2007)), there is a lack of research in the humanitarian context. To address this gap, we have embarked on a study to investigate the key facets of co-design expertise in HE and how co-design expertise can be assessed and developed in engineering degrees and courses. As a first step in this research project, this paper focuses on answering the following research question: how do experienced humanitarian engineers' co-design strategies differ from those of novices? To answer this question, we took a scenario-based assessment approach similar to that used in other studies (e.g., Adams, Beltz et al. (2010), Atman, Adams et al. (2007), Kilgore, Atman et al. (2007), McKenna (2007), McKenna, Hynes et al. (2016)). Specifically, we asked first year engineering students and experienced professionals to complete a scenario-based instrument, the Energy Conversion Playground (ECP) design task (Mazzurco, Huff et al. 2014), and then developed a rubric assessing co-design expertise based on the participants' responses and literature on the topic. Outcomes from this study will enable educators to use scenario-based assessment in their own research, teaching, and curriculum development to promote co-design in humanitarian engineering.

Literature review

In this review, we synthesise literature on two topics that form the basis of our study: design and community participation in humanitarian engineering, and scenario-based assessment.

Design processes and community participation

Design of appropriate technology is at the core of humanitarian engineering practice. Key aspects of appropriate technologies include that they must match the socio-cultural and economic realities of the beneficiary communities, be designed following a rigorous process, and appropriately involve community members throughout the process (Lucena and Schneider 2008, Schneider, Leydens et al. 2008, Mattson and Wood 2014). Consequently, HE scholars and practitioners have invested extensive efforts in trying to conceptualise design processes and community participation in HE projects.

In terms of design process, multiple models have been proposed that share many similarities and some differences. For instance, Sianipar, Yudoko et al. (2013) propose a design methodology for appropriate technologies that includes four stages. It starts with the planning stage, focused on understanding stakeholders' requirements, and finishes with the assessing stage, focused on evaluating the developed technology. Others have included more steps, unpacking the design process in greater detail (e.g., Drain, Shekar et al. (2017), Engineers without Borders Australia (2019), Murcott (2007), Sianipar, Yudoko et al. (2013), Ssozi-Mugarura, Blake et al. (2017)). Nonetheless, all design processes cover more or less the same range of activities, although providing different names. This can be seen clearly in Figure 1, where these different processes have been set out such that vertically aligned design phases cover similar activities. The only exceptions are perhaps Murcott (2007) and Drain, Shekar et al. (2017), who add an implementation stage that goes beyond the last stage of other proposed design processes. While the processes presented in Figure 1 were specifically developed for HE or similar projects, they also align with traditional design processes, such as the double diamond, which includes four phases: discover, define, develop, and deliver (Design Council 2019).

[Figure 1 here]

Another difference among the processes in Figure 1 is in the language used by Ssozi-Mugarura, Blake et al. (2017) and Murcott (2007). They use terms such as *collaborative* design, problem *co*-definition, idea *co*-generation, and others, that emphasise the importance of involving users across the lifespan of a design project. This emphasis on community participation reflects the first of the nine principles identified by Mattson and Wood (2014), which states: "Co-design with people from the specific developing-world context encourages designer empathy, promotes user ownership, and empowers resource-poor individuals" (p. 121403-2).

This concept of community participation, however, is very complex and requires further unpacking. Mazzurco, Leydens et al. (2018) propose a three-level framework to understand community participation in HE. The bottom level is *passive*, in which the community is not involved, followed by *consultative* and *co-constructive* levels, which reflect increasing levels of participation. Building upon the work of the World Bank Popular Participation Learning group, Engineers without Borders Australia (2018) identify four levels of participation starting at the *information-sharing* level in which people are told about a project and ending at the *initiating-action* level in which people organise themselves to initiate and complete the project. Furthermore, these frameworks are not unique to HE and, in fact, align with, and often build upon, participation models beyond HE.

In design more broadly, Sanders and Stappers (2014) proposed two participation levels: *design for* and *design with*. In the former, users are seen as subjects of the design process, whereas in the latter they are seen as partners. Likewise, Druin (2002) proposed a model of children's participation in technology development, suggesting that children can take the role of *users* of a completed technology, *testers, informants,* or *design partners*, in increasing order of participation. Similar frameworks can be found in international development. For instance, Kanji and Greenwood (2001) used a five-level participation framework ranging from *compliance,* in which community outsiders assign tasks to locals but withhold decision-making power, to *collective action*, in which locals set their own agenda and conduct the project independently.

Figure 2 depicts these participation frameworks. Participation levels have been aligned horizontally to reflect similar degrees of community participation. Note that whilst lower levels are not usually preferred, higher levels of participation may also not be ideal if forced upon the community (Cooke and Kothari 2001, Mazzurco, Leydens et al. 2018).

[Figure 2 somewhere here]

Scenario-based assessment

There has been a long debate on how to assess the wide range of skills that engineers need for practice (Shuman, Besterfield-Sacre et al. 2005). In engineering education, self-report instruments, like Likert scales, have often been used to assess a large number of constructs, ranging from environmental knowledge and design skills, to generic engineering competencies (e.g., Carberry, Lee et al. (2010), Azapagic, Perdan et al. (2005), Peeters, Londers et al. (2014)). However, while self-report instruments can be administered to large number of participants, they have been often criticised for not being able to directly assess behaviours (Peng, Nisbett et al. 1997).

In other disciplines, simulation-based assessments (Ilgen, Ma et al. 2015) and assessment centres (Hoffman, Kennedy et al. 2015) have often been used to assess behaviours because they evaluate performance in realistic situations. The downside is that they are hard to scale and therefore only used with small numbers of participants.

A third alternative is scenario-based assessment, which involves asking participants to respond to questions related to a short case-study (or scenario). This makes scenario-based instruments typically more time-consuming to score than self-report scales, but arguably better at evaluating the potential behaviours of participants in realistic situations (McKenna 2007), while being more easily administered to larger numbers of participants than assessment centres (which are quite labour-intensive forms of assessment). Because of these characteristics, scenario-based instruments have been used to assess multiple engineering-related competencies, including design skills (Atman, Adams et al. 2007), interdisciplinary problem-solving (Adams, Beltz et al. 2010), and knowledge of global, societal, economic, and environmental contexts (McKenna, Hynes et al. 2016).

Scenario-based instruments are comprised of three elements: a scenario, a set of questions related to the scenario, and a scoring system. The length of scenarios can range from one sentence (Kilgore, Atman et al. 2007) to multiple paragraphs long

(McMartin, McKenna et al. 2000). No matter the length, the most important feature of the scenario is its authenticity. They must describe a realistic situation. Therefore, scenarios are usually inspired by case-studies of real practice (Borenstein, Drake et al. 2010)), practitioner interviews (Thoma, Derryberry et al. 2013), or a combination of the two (Jesiek, Woo et al. under review).

Scenario-based instruments can use questions that are either closed- or open-ended. Responses can be captured through think-aloud protocols (e.g., Atman, Adams et al. (2007)) or online or paper-based forms (e.g., Adams, Beltz et al. (2010), McKenna, Hynes et al. (2016)). Responses to close-ended questions are evaluated by computing scores and analysed using statistical methods (e.g., Jesiek, Woo et al. (under review)), whereas responses to open-ended questions are coded and analysed using rubrics (e.g., Kilgore, Atman et al. (2007), McMartin, McKenna et al. (2000)). For open-ended questions, developing the rubrics is the most critical and time-consuming part and it usually requires two sequential steps. First, at least two researchers code the responses independently and then meet to compare coding and develop a rubric. Second, the rubric is tested. At least two researchers use the rubric to score a sub-set of responses, meet to reach consensus, and then iterate this process. The consensus building process allows also refining and improving the rubric. Agreement between the researchers usually starts low (e.g., 50% to 75% agreement (Atman, Adams et al. 2007, Hess, Beever et al. 2014)) and then improves over time (85% to 97% agreement (Hess, Beever et al. 2014)). Having at least two researchers working independently throughout the entire process helps mitigate any biases. However, it requires raters to undergo multiple iterations of coding before obtaining strong inter-rater reliability, thus making the process more time-consuming.

In sum, scenario-based instruments can take many forms and can be used to reliably assess multiple constructs. The choice of question type informs their ability to capture nuances and their scalability. Instruments with open-ended questions are more time-consuming to analyse and require extensive work to reach high level of reliability, but are able to capture more subtle nuances in the data than instruments with close-ended questions that are more easily scalable.

Methods

This paper reports on one aspect of a larger study on the further development, validation, and application of the Energy Conversion Playground (ECP) design task (Mazzurco, Huff et al. 2014). The purpose of the ECP design task was to offer an easily administered, reliable, and valid measure of socio-technical thinking and co-design expertise in the context of humanitarian engineering (HE).

This paper reports on the extent to which the ECP task can reliably assess differences in co-design expertise between novices (students) and experts (engineering academics and practitioners with experience in HE). Other parallel publications will report on the part focused on socio-technical thinking (Mazzurco and Daniel in press), and on pre-post (Mazzurco, Daniel et al. 2019) and longitudinal studies characterising any development of expertise over time or due to targeted educational interventions.

The ECP design task

The first versions of the ECP design task were developed in previous studies (Mazzurco, Huff et al. 2014). These studies demonstrated that the ECP design task was sensitive to changes in socio-technical thinking in students who had participated in workshops on engineering for community development. The same scenario used in these initial studies has been used again here:

> In developing countries, energy production is one of the most critical problems. Resources or technologies to produce energy are often not available. Thus, human power conversion systems might be used to power small appliances. Imagine that you and your team are assigned to a design project in partnership with a Non-Governmental Organization (NGO) of a developing country. The NGO needs a low-cost power system that can generate enough energy for the lights of a primary school. One of the members of your team suggests using merry-go-round, seesaw, and swing to produce energy that can be converted to electricity for the lights. (Mazzurco, Huff et al. 2014)

The scenario was adapted from a conference paper by Pandian (2004), which focused solely on the technical details of how playground equipment (such as merry-go-rounds, seesaws, etc.) could be adapted to create energy to power small appliances. Another inspiration was the PlayPump failure, in which children spinning on merry-go-rounds would provide the energy to pump bore water into a reservoir. The PlayPump project is often used as a salutary example of how well-meaning projects can fail without a deep understanding of the social dynamics and broader contextual issues that underpin HE projects (Borland 2011). Therefore, creating the ECP scenario based on these two examples gives it authenticity, as well as enough open-endedness to capture different types of thinking.

After the above scenario was given, respondents were asked two open-ended questions:

1. What considerations do you need to take into account to solve the problem described in the scenario? List and describe all constraints and justify their inclusion.
2. How would you proceed to solve the problem described in the scenario? List and describe concisely all the steps you would take to solve the problem described in the scenario

Question 1 was used in the original study (Mazzurco, Huff et al. 2014), whereas Question 2 was asked only in this present study. While Question 1 relates to socio-technical considerations, Question 2 was designed to elicit respondents' conceptions of humanitarian design processes. To ensure that responses represented respondents' own understanding, they were not given definitions of socio-technical thinking or co-design prior to completing the scenario. Although the scenario and both questions were distributed to all participants in this study, this paper is focused on analysing Question 2 responses, on co-design.

Context of data collection

The goal of the recruitment process was to maximise the diversity of respondents' background to so ensure a range of co-design expertise. Similar to other novice-expert studies (e.g., Atman, Adams et al. (2007)), we used purposeful sampling to recruit participants from two very contrasting groups: engineering students at the very beginning of their degree (n=26), and engineering practitioners (n=16). The students were taken as 'novices' as, although they align themselves with engineering, they were yet to have any formal training. We did not have any means to objectively determine whether all practitioners could be considered 'experts' prior to recruitment. However, by targeting engineers with a range of experiences in humanitarian engineering (described below), we maximised the probability of including participants with high levels of expertise. The participant numbers are similar to other comparable studies (e.g. Atman, Adams et al. (2007)).

Student recruitment

The participating students were part of the first cohort of a new practice-based engineering degree at an Australian university. Their responses were collected during a workshop in the orientation week, prior to the start of their very first engineering unit. Of the 28 students in attendance, 26 returned completed responses to the scenario-based instrument. Students completed the scenario a second time, at the conclusion of the first unit of study in their degree. This unit was focused on human-centered design, and lead students through the Engineers without Borders (EWB) Challenge (Jolly, Crosthwaite et al. 2011). Twenty-one students returned completed responses to this post-unit administration of the instrument. Therefore, we used a total of 47 student responses in developing our rubric. However, the statistical analysis of the pre vs post changes is not included in this manuscript as it has been already reported elsewhere (Mazzurco, Daniel et al. 2019).

Together with students' responses to the scenario, we also collected demographic data. The goal of collecting demographic data was not to use it to make claims about how demographics impacted the responses (as the sample was too small), but rather to provide some background to the key characteristics of the student group that we recruited. Twenty-two students returned completed demographics responses. The demographics of our sample are similar to the population of first-year engineering students in Australia (Engineers Australia 2018), with 16 who identified as male, five as female, and one as non-binary, and being mostly school-leavers of European descent. Following the approved ethics protocol, the collected responses were anonymised by a research assistant to protect the identity of the respondents. Students were assigned a three-letter code, followed by either a 1 or a 2 to indicate a response from either the pre- or post-test respectively.

Practitioner recruitment

Practitioners were recruited through two means: the Humanitarian Engineering Education of Australasia Network (HEENA) and the Journal of Humanitarian Engineering (JHE). The HEENA is an informal group of academics and professionals interested in HE, while the JHE is an open-access peer-reviewed journal with the aim of improving best practice in HE. Participants were recruited through the HEENA

Facebook page and invited to complete the scenario and demographics questionnaire online via Qualtrics. The same invitation was sent to all past JHE authors. We selected these two groups because of the alignment between their background and experience and the HE focus of this study. A total of 26 people opened the questionnaire, of which 16 provided complete responses. However, one of the respondents was dropped as they provided only an anecdotal account of their experience on another project, rather than responding to the actual question. The remaining 15 respondents were assigned a one-letter alphabetical code, ranging from A to P.

The demographic data of the participants indicates we were successful in recruiting participants with a wide range of backgrounds, aligned with our aim of characterising a wide spectrum of co-design expertise. Table 1 reports a summary of the practitioners' demographics and experiences.

[Table 1 here]

Data Analysis

The data analysis was comprised of two stages – qualitative interpretation and analysis of responses to develop a rubric, then quantitative analysis of rubric scores. That is, first we developed a rubric to characterise co-design expertise through an iterative and inductive process of cycles of coding and thematic analysis of the pool of responses (15 from practitioners and 47 from students). The rubric was also sense-checked against the literature on co-design in humanitarian engineering and similar contexts. Then second, we compared scores against this final rubric of students and practitioners using descriptive statistics.

In interpretivist research, it is important to understand the worldview of the researchers involved. Although we grounded our analysis of responses in the literature, how we made sense of this was necessarily through our own lenses. Therefore, before describing our process in developing the rubric, we briefly describe our backgrounds and perspectives.

Both of us have lived and worked in multiple countries. We are both motivated to work in humanitarian engineering because of the incremental way in which humanitarian engineering education can contribute to social justice. Both of us are scholars in engineering education, with expertise in interpretivist research. Our experience in humanitarian engineering slightly differs. The first author has primarily worked with Engineers without Borders Australia, on overseas study tours and in research through the Journal of Humanitarian Engineering. The second author's PhD research was on humanitarian engineering (Mazzurco 2016), and he is an experienced facilitator with both the EWB Challenge and the EPICS program (Coyle, Jamieson et al. 2005).

The qualitative process to develop the final rubric that we present in the results was highly iterative. As a starting point, we segmented students' responses into individual statements, each representing one proposed problem-solving step. Each author analysed these segments independently and then met to discuss their analysis. This led us to develop a rubric with 11 categories representing design steps. However, when we attempted to apply this first iteration of the rubric to the practitioner data, several issues became apparent. Foremost of these was that while the rubric was complex, the scores it generated seemed at odds with our intuition and the extant literature about what constituted an expert response.

As a consequence, we decided our next step would be to consider the responses holistically. We independently coded complete responses as either 'excellent', 'good', or 'not good', based on our own understanding of best practice having each taught HE design for several years. Then we met to discuss our agreements and identify which aspects of the responses made them ‘excellent’, and how this compared to the relevant literature. In the first iteration of the rubric, ‘involvement of the community’ had been one of the 11 categories. However, we now recognised this as an important discriminant of expertise. Instead of having it simply alongside the other 10 categories, each representing a different step in the design process, we reframed it as an orthogonal dimension of quality: in this next iteration of the rubric, different design steps would be scored for the extent to which they described involving the community. This was scored on three tiers: ‘not mentioning community’, ‘information transfer’, and ‘collaborative’, in line with other characterisations of community involvement (e.g., Mazzurco, Leydens et al. (2018)). Furthermore, the ten categories we had originally developed to represent different aspects of the design process were collapsed to four design phases: *scoping, concept development, detailed design,* and *delivery*, well-aligned with some of the literature cited earlier (e.g., Design Council (2019), Engineers without Borders Australia (2019)). This was done in recognition of the tacit expertise of experienced practitioners — quality responses need not spell out in detail all ten steps we had identified.

We independently applied this rubric to the 11 practitioner responses and then met again to reconcile our disagreements and reach consensus. We agreed that the rubric now could successfully differentiate various levels of expertise. However, while the *scope* and *delivery* phases were easy to distinguish, there was ambiguity around the distinction between *concept development* and *detailed design*. In what became the final version of the rubric, we decided that making this distinction added complexity without adding insight, and so we collapsed these together as simply *develop*.

We evaluated this decision by independently applying this version of the rubric to the 4 remaining practitioner responses that had come in late. When we met to discuss our analysis and reach consensus, we agreed that combining *concept development* and *detailed design* did not lose any important detail and still distinguished quality responses. The final stage of our analysis was to independently apply this rubric to the set of student responses. We then met together to discuss our disagreements and reach consensus.

There were no further changes to the rubric. Its final form is reported in the next section, along with descriptive statistics reporting our results from applying it to the student and practitioner responses.

Validity and reliability

Bernhard (2018), in his editorial, reflects on the maturation of engineering education research and its relevance for practice, and how this is contingent upon research quality. We addressed this in both our data collection and analysis, by being deliberate about ensuring the reliability and validity of our approach. We used the 'Qualifying Qualitative Research Quality' framework developed by Walther, Sochacka et al. (2013), which outlines different aspects of quality in both the 'making data' (i.e. data collection) and 'handling data' (i.e. data analysis) stages: procedural validation, pragmatic validation, ethical validation, theoretical validation, communicative validation, and process reliability.

A full description and analysis of how we applied this framework will be the subject of future publication, while here we give some brief examples. In terms of making our data (i.e. designing an assessment, recruiting participants and collecting their responses), we devised an authentic scenario adapted from real studies (cf. procedural validation). We used purposeful sampling to recruit participants with either no HE experience or training, or demonstrable HE experience and training, to validate our comparisons of novices and experienced practitioners (cf. pragmatic validation).

This method of studying expertise follows what Chi (2006) termed a relative, as opposed to absolute (i.e. studying exceptional people alone), approach (cf. theoretical validation). Participants' responses were de-identified by a research assistant, and student responses were only assessed after grades had been finalised. This was communicated to participants, according to our approved ethics protocol, to empower them to answer freely (cf. ethical validation).

In handling the data (i.e. developing and applying a rubric for co-design), we cross-checked successive iterations of our rubric against the literature on co-design (cf. theoretical validation). For example, both our design phases and our participation levels can be easily mapped against existing frameworks, as we demonstrate in more detail in the discussion. We always coded responses independently before meeting together to reach consensus (cf. process reliability), and have been transparent about describing our own worldviews to help the reader make sense of our interpretation (cf. communicative validation). To further address communicative validation, we presented our rubric and analysis in workshops to final-year research students and academics in humanitarian engineering to ensure it made sense to relevant audiences.

As an additional test of reliability, we asked two engineering education scholars with experience in humanitarian engineering to attend a sense-checking workshop. During this 1-hour workshop, we gave them an overview of the rubric and then asked them to rate five responses using the rubric. Once they had rated the responses independently, we asked them to compare their responses and come to consensus. Finally, we shared our ratings and discussed differences. In the discussion, both raters observed feeling they needed more time to become familiar with the rubric and that they tended to be less conservative than us in their interpretation of the rubric. That is, they generally scored responses higher.

This is reflected in Table 2, which reports percentage of agreement, correlations between responses, and the gradient of the line of best fit between responses. The mediocre percentage agreements and the raters' observation that they needed more time suggest that the rubric is a valid but hard-to-use instrument requiring extensive training. Nonetheless, it interesting to note that the correlations increased with the consensus-building process (i.e., the raters' final consensus was more correlated to our consensus than their original ratings).

In addition to the correlation, it is also interesting to consider the gradient of the line of best fit. Ideally, this gradient would be 1, indicating that the step-size for increases in response quality is assessed equally by both raters. The gradient of 0.78 between our consensus scores and the raters' consensus scores indicates they were more liberal in scoring responses higher. This suggests that more training on the rubric is required, emphasising the importance of relying only on what actually is said in the responses, and avoiding any generous, idiosyncratic over-interpretation. This conclusion aligns with other studies cited in the literature review, showing that raters usually require multiple iterations before achieving good levels of agreement.

[Table 2 near here]

Results

The results are presented in two main sections. In the first section we detail the rubric we developed, with discussion of exemplars and boundary cases to clarify the distinctions between different rubric criteria, and so provide insight into differentiating different levels of expertise. In the second section, we report our quantitative analysis of applying the rubric to the students' and practitioners' responses.

Rubric overview

In simple terms, the final version of our rubric has two aspects: design phases and community participation. Our rubric recognised three design phases, each of which is given a quality score from 0-3 depending on the level of community participation. These three scores are not intended to be summed, but instead reported as three distinct dimensions to preserve a profile of how community participation is described in different phases of the design process.

The three design phases are *scope, develop,* and *deliver*. The *scope* phase focuses on understanding the problem, context, and stakeholders, to so clarify the project requirements and goals. The *develop* phase includes both conceptual and detailed design activities, whereas the *delivery* phase focuses on solution implementation and long-term sustainability. This delineation of the design process is consistent with the literature on design and was validated in our analysis by how rarely there was any ambiguity about categorising particular statements into these different design phases. Typical activities associated with each phase are given in the rubric in Table 3.

[Table 3 near here]

The rubric in Table 3 also describes how these different design phases were scored. Responses were scored by whether or not they included activities belonging to the different phases, and how they described community participation. If responses did not include any mention of any activities typical of, say, *scoping*, they received a score of 0 for that design phase. Conversely, if they mentioned at least one activity belonging to a design phase, but without mentioning any form of engagement with stakeholders, they received a score of 1. For example, respondent H wrote “Determine how much energy is realistically needed” and so was awarded a score of 1 for *scoping*.

Scores above 1 were reserved for those responses that included some mention of stakeholders. If this was only at the level of *information transfer* (e.g. respondent I: "Interview the school employees to get the typical work-day schedule and needs"), they received a score of 2 for that design phase. On the other hand, if in at least one part of their response they described *collaborating* with community members, this was awarded a 3. For example, respondent B wrote "engage in a participatory, iterative brainstorming process with the community" and so was scored a 3 for *develop*.

Exemplar excerpts for each design phase and community participation level are provided in Table 4. For example, in the top left, an excerpt is given from respondent AJX1: "Assess how much energy the school will need". This has been interpreted as 'identifying design requirements', but without any mention of stakeholders, and so has been scored a 1 for merely mentioning *scoping*. The excerpt in the centre of the table, from respondent I, describes brainstorming different ideas, and selecting between them with advice from stakeholders. As such customer feedback represents a flow of information between the engineers and stakeholders, this excerpt was scored 2 for *develop*. The excerpt from respondent O in the bottom right represents an expert-level *collaborative delivery*. It describes putting the detailed design out to tender in the local community, and collaborating with local personnel for the "construction, operation, and maintenance" of the design solution.

[Table 4 somewhere here]

These excerpts above were selected as they each are representative examples of how we characterised different levels of community participation across the three design phases. Conversely, in the following section, we describe boundary cases. These were responses that did not fit neatly within the rubric cells. Our intention in discussing how we eventually coded them is to further illuminate the distinctions we drew between different levels of community participation and potentially assist the reader in using the rubric in a similar way with their own data.

Boundary cases

Although the above exemplar responses were straightforward to code, with some responses this was more difficult, having some ambiguity in how they could be interpreted – so called 'boundary cases'. These boundary cases highlighted three key issues in using the rubric:

- Not assuming what is not stated explicitly
- The importance of key words
- Interpreting the whole response

These are explored in detail below.

Not assuming what is not stated explicitly

In our analysis discussions, we had to frequently guard against assuming things the respondents had not explicitly stated. These discussions highlighted the importance of blinding ourselves to the expertise of the respondents. That is, knowing a particular ambiguous response came from an experienced practitioner might bias us in being more generous in its interpretation, or more critical if we knew it came from a novice student. To overcome this potential bias, we always returned only to what the respondent had stated explicitly, and limited how many inferences we drew about different ways the step they described could be enacted.

For example, respondent M stated in part "Conduct needs assessment (with all stakeholders including the school, teachers, students, NGO)". This *scoping* activity could conceivably be done in a *collaborative* manner, but also simply by surveys or one-on-one interviews – i.e. *information transfer*. Because of this ambiguity, we gave a score of 2 as *collaboration* was not described explicitly.

Other examples included where stakeholders or the community were mentioned, but with ambiguity as to whether there was any intended interaction at all. *Collaboration*, or *information transfer*, had to be made explicit. For example, respondent ARW2 mentioned "create a stakeholder group". This could be interpreted generously as hosting a focus group, collaborating with community representatives, or alternatively could simply mean creating a list of who the stakeholders are. Seeing as the only other mention of the community by this respondent was in the sense of educating them (i.e. *information transfer*), the comment "create a stakeholder group" was interpreted only as mentioning *scoping* – and so scored a 1 for this design phase. Whenever there was such ambiguity, we took the most conservative interpretation. That is, community participation had to be mentioned explicitly to score anything higher than a 1.

Only one response had ambiguity about the design phase. This came from respondent RHL2, who wrote "Bring finished report and ideas to contractors and discuss". This statement bridged the *develop* and *deliver* phases, as the "finished report" could represent the endpoint of the *develop* phase, whereas engaging contractors could be the first step of the *deliver* phase – as a precursor to building the design. With the caveat that contractors are not necessarily community members, we scored this as 1 for mentioning both *develop* and *deliver*.

The importance of key words and phrases

In our assessment of community involvement, the inclusion of a key word or phrase was pivotal to distinguish between levels. For example, respondent G wrote in part "Co-design options. Present and get feedback. Iterate. Design, develop." This was scored 3 for *develop*. We recognised however that without the prefix “Co” of “Co-design”, this would have been only scored as *information transfer* because of the sentence “present and get feedback”. That is, our scoring hinged on the inclusion of the key prefix “Co”, without which our interpretation would have been very different.

Another example comes from respondent B, whose full response was:

> 1. Meet with community leaders (gatekeepers) to better understand the issue(s) that are important in the community
> 2. If the energy production priority aligns with community priorities, gather a roundtable of community stakeholders- particularly voices who may not have been previously heard (women, children)
> 3. engage in a participatory, iterative brainstorming process with the community
> 4. with the community, decide on which solution best meets the community needs
> 5. with the community, implement the solution

Without the words “participatory” and “with the community” being included in the latter three statements, they would have only been scored as 1 for merely mentioning the *develop* and *delivery* phases. As is, however, these statements were scored as *collaborative development* and *delivery*. Any ambiguity in whether “with the community” could mean *information transfer* versus *collaboration* was resolved by how explicitly participation and collaboration were described in the first two statements. That is, responses were judged holistically.

Interpreting the whole response

Respondent B was judged as *collaborative development* and *delivery* because the repeated use of the phrase “with the community” was interpreted as *collaboration*, as it was preceded by phrases such as “gather a roundtable of community stakeholders”. That is to say, the whole response was used to help choose between competing interpretations of unclear statements.

Other responses also demonstrated the importance of interpreting the whole response, rather than just piecemeal individual statements. An example comes from respondent DJX2:

> Research the problem
> Consult the community
> Create a refined problem statement
> Establish constraints & set budget
> Report back to community & make adjustments if needed [consistently throughout project]
> brainstorm ideas → test with prototypes
> refine ideas so they can be implemented easily

This response would have been scored only a 1 for *develop* and *delivery* (mentioning only the statements "brainstorm ideas → test with prototypes" and "refine ideas so they can be implemented easily", for these two design phases respectively), were it not for the phrase "consistently throughout the project" that the respondent wrote next to the sentence "Report back to community and make adjustments if needed". Because of this key phrase, DJX2 was scored as *information transfer* for all 3 design phases. Another potentially ambiguous statement comes from respondent BTT1: "Build it with locals", highlighted in italics in the full response below:

> Introduce the project to the community
> - what's going to happen
> - why it's going to happen
> - who
> - when
> - where
> - how
>
> Educate the community on how it works before beginning production
> Prototype the project to ensure it will be correct
> *Build it with locals*
> Ensure that it runs as it's supposed to and that the locals know how
> Educate the community that if it's to break then they know how to fix it

Although this could conceivably be done collaboratively, interaction in the rest of the response was only ever characterised by one-way information transfer from the engineer to the community (e.g. "Introduce the project to the community: what's going to happen, why it's going to happen, who, when, where, how"), and so "build it with locals" was interpreted as *information transfer* for *delivery*. That is, in general any ambiguity was resolved by interpreting statements in the context of the whole response.

Analysis of responses

In this second part of the results, we report our quantitative evaluation of student and practitioner responses. First, we use graphical representations to visualise the key differences between individual and aggregated student and practitioner responses. Second, we report the distribution of scores across the two dimensions of design phases and community engagement, to compare trends between novices and experienced practitioners.

Graphical representations of co-design expertise

Figure 3 below depicts our analysis of four contrasting responses. We represent co-design expertise on two dimensions – successive design phases on the horizontal axis, and depth of community engagement on the vertical axis. This signifies that, although the design process is iterative, there is a progression from *scoping* problems to *developing* design ideas, then *delivering* (i.e. building and maintaining) solutions, which does not usually occur in the reverse direction when designing new products.

Each of the responses portrayed in Figure 3 has been discussed to some extent earlier. Respondent B was one of only two that described a *collaborative* approach to each of the three design phases, repeating the key phrase "with the community", and so epitomises an expert approach. Respondent DJX2 described taking an *information transfer* approach throughout the design process, as indicated by the key statement "report back to community and make adjustments if needed - consistently throughout the project". Respondent L (see Table 4) described *scoping* in detail and, apart from assessing to what extent they would be possible (i.e. *scoping*), did not mention any aspects of the *develop* and *deliver* phases. Finally, respondent AJX1 (see Table 4), a pre-unit student response, represents a typical novice approach to the scenario – merely *mentioning* aspects of *scoping* and *develop*, but without any description of community involvement.

[Figure 3 here]

In Figure 4, the median co-design scores have been plotted for all students (i.e. combined pre- and post-) versus practitioners. Practitioners typically described greater community involvement than students, except in the final design phase *deliver*. In the following section, we go into more detail about the distribution of scores between students and practitioners.

[Figure 4 here]

Analysing the distribution of responses between novices and practitioners

In Table 5, we have compared the distribution of scores for the different design phases for novices and experts. To this point we have cited examples from across the entire data set (i.e. pre- and post-intervention student responses, as well as practitioner responses) to demonstrate different aspects of the rubric, as we used this whole data set to generate the rubric. However, in the analysis below of novices versus the experienced professionals, we present a comparison of results between practitioners (i.e. experts, n = 15) and students' pre-unit responses (i.e. complete novices, before any exposure to principles of human-centred design, n = 26). The analysis of students' pre- and post-intervention responses, to so examine whether participation in a course of study in human-centred design had any effect discernible by our instrument on socio-technical or co-design expertise, is the subject of another publication (Mazzurco, Daniel et al. 2019).

[Table 5 here]

There are several important observations to make from this table. The key observation is that not one of the 26 students described a collaborative approach to either the *scope, develop,* or *deliver* design phases. Another interesting observation is that a sizeable percentage of the practitioners scored 0 or 1 for each of the design phases. These observations will be discussed in the following section.

Discussion

In this section, we discuss our findings and relate them to the literature on co-design, how educators can use our instrument, and some limitations of our work and the future research required to address them.

Comparing our findings to the literature

From our analysis, we characterised the design process as having three phases: *scoping, develop,* and *deliver*, where these can take place with different levels of community involvement, from not acknowledging the community at all, to collaborating as peers. How does this compare with the literature on participatory design?

In Figure 5, we have mapped different characterisations of the design process to ours. One interesting observation is that although we accorded *deliver* the same significance as *scope* and *develop* (as indeed a project cannot have any success unless it is delivered), it is seemingly under-represented in the design literature. It is perhaps then no coincidence that practitioners scored lowest on this design phase (cf. Figure 4). We argue therefore, as do Russell and Vinsel (2019), that considerations of delivery, such as maintenance, should figure more prominently in engineering education.

[Figure 5 here somewhere – on its own landscape page]

In Figure 6, we have mapped different characterisations of community engagement to ours. Whereas the top level could be characterised as participatory co-design, the 2nd level can be summarised as human-centred design. Although we characterised participatory co-design as expertise, this is not to suggest that any activity that is not participatory is not of value, nor that every activity should always be participatory. For example, secondary research certainly has value in design, even though of course it involves no community participation (Mazzurco, Leydens et al. 2018). The concern would be if the only *scoping* that took place was secondary research.

[Figure 6 somewhere here]

Conversely, with the logistical constraints of a real project, community participation time may be better spent on some design processes than others. For example, there may be much more value in collaboratively unpacking the relative importance of different selection criteria and so developing a decision matrix to evaluate different design ideas, rather than collaboratively brainstorming those new ideas in the first place, or collectively developing a detailed design. There is also the danger that pushing for community participation can verge on coercion, and abuse of a power dynamic (Cooke and Kothari 2001). Nonetheless, an expert co-designer would always try to involve community members as peers.

Novice versus experts

Our aim was to characterise the development of co-design expertise, and to that end we recruited students as 'novices', and experienced practitioners as people more likely to have high levels of expertise. None of the 26 pre-unit student responses described collaboration in any design phase (cf. Table 5), and in fact their median score across the three phases was 1 – i.e. the typical student response described different aspects of design, but without mentioning any community involvement. Our assumption that we could use pre-instruction students as ‘novices’ is therefore justified. The corollary is that our instrument is capable of discriminating expertise, or the lack thereof.

However, given the practitioners’ low median scores, it seems the practitioners we recruited were perhaps not all the most exemplary co-design experts. Another interesting observation is how low practitioners scored for *deliver*– fully two-thirds did not mention any stakeholders whatsoever in describing *delivery*. Whether this is because collaboration is not as important in this design phase as it is in *scoping* and *develop*, whether it’s an artefact of our instrument (i.e. that our instrument does not prompt responses about *collaborative delivery*), or alternatively that the practitioners we recruited were not all experts, is an open question. What is important, however, is that *some* of the practitioners were experts, and that by collecting a range of responses, from both students and practitioners, we have nevertheless been able to develop a rubric for characterising different levels of expertise.

Implications for educators

The main contribution of this work is the development of our rubric for co-design, which we hope other educators and researchers will be able to use in their own work. For example, researchers can use it to make pre/post comparisons of co-design expertise to evaluate different educational interventions, longitudinally to track the development of expertise over time, or as a one-off assessment of expertise.

However, we also see potential for using the instrument as a teaching tool. The instrument itself can act as a discussion starter, but there are also possibilities in using the rubric for peer-assessment. By training students in discriminating different levels of expertise in the rubric, and then using that to evaluate the responses of their peers, this offers another perspective for students to consider and reflect on the design process and different levels of community participation.

More broadly, our findings also provide insights for educators and institutions creating new humanitarian engineering courses or degrees. First, such programs should emphasise all aspects of the design process. Second, students should be given opportunities to explore and critically examine ways to involve community in all stages of the design process.

Limitations and future research

There are several limitations to our study. Although our sample size was appropriate for characterising expertise and developing a reliable rubric, it is too small to make any generalisable claims regarding differences of co-design expertise based on demographic data. For instance, the women in our sample tended to score higher across all design phases than the men. However, this finding cannot be generalised and could just be an anomaly of our data. In the next iterations of data collection, we will collect larger samples of data to investigate differences based on demographic and experiential variables.

With all instruments involving open-ended responses, there is the potential for ambiguity in interpretation. We have sought to clearly delineate our scoring rubric by describing both exemplars and boundary cases for the different criteria, but one test of how clearly our rubric discriminates different levels of expertise will be whether other researchers can use the instrument with a high-degree of inter-rater reliability with ourselves and so ensure valid comparisons between cohorts.

We designed our rubric to be simple, having only a 3-point scale across the 3 design phases. In exchange for simplicity, we have foregone nuance. Using the instrument with larger samples, especially in pre-/post- or longitudinal study designs, will help us evaluate if our rubric is sufficiently granular, or too coarse, to evaluate small developments in expertise. Like for other scenario-based instruments, another challenge of the rubric that emerged during our sense-checking workshop is that it takes some time to use it accurately and extensive training is needed. Therefore, we would invite researchers wanting to use the rubric to contact the authors to discuss appropriate training opportunities.

Another limitation is that our scenario is only in the context of energy. Although the literature suggests that the design process does not vary in different HE sectors, perhaps our instrument does not capture nuances unique to other sectors, such as water, sanitation, and hygiene (WASH), or agriculture. In future, we hope to develop similar instruments in these other HE sectors, and cross-validate them against each other as assessments of co-design expertise in humanitarian engineering.

Finally, one question with all such scenario-based assessments is how well do they actually predict real behaviour in practice. Although scenario-based instruments arguably are better at predicting the performance of participants in realistic situations than self-reports (McKenna 2007), an area of future research would be evaluating the validity of these instruments in predicting approaches to co-design in authentic humanitarian engineering contexts.

Conclusion

Through analysing responses from students and experienced practitioners to a scenario-based assessment, we have developed a rubric to assess co-design expertise in the context of humanitarian engineering. By detailing our data collection and analysis processes, and by comparing novice and expert responses, we have validated it as a measure of expertise. The rubric characterises co-design as involving three phases: *scoping, develop,* and *deliver*, and evaluates each of these phases using the quality criterion of the extent to which community involvement is described. The rubric has been described in detail, so that readers can adapt it in their own research or use it as a teaching tool to prompt discussions or for formative assessment.

Acknowledgements

The authors wish to thank the respondents in this study, those students and experienced practitioners who volunteered their time to participate. We also wish to thank the members of the research group in the XXX at XXX who participated in a workshop to establish the communicative validity of the research. Their feedback helped improve the clarity of the rubric. Finally, thank you to the anonymous reviewers whose constructive feedback helped improve the quality of the paper.

References

Adams, R. S., N. Beltz, L. Mann and D. Wilson (2010). "Exploring student differences in formulating cross-disciplinary sustainability problems." International Journal of Engineering Education **26**(2): 324.

Atman, C. J., R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg and J. Saleem (2007). "Engineering design processes: A comparison of students and expert practitioners." Journal of engineering education **96**(4): 359-379.

Azapagic, A., S. Perdan and D. Shallcross (2005). "How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum." European Journal of Engineering Education **30**(1): 1-19.

Bernhard, J. (2018). "Engineering Education Research in Europe – coming of age." European Journal of Engineering Education **43**(2): 167-170.

Borenstein, J., M. J. Drake, R. Kirkman and J. L. Swann (2010). "The Engineering and Science Issues Test (ESIT): a discipline-specific approach to assessing moral judgment." Sci Eng Ethics **16**(2): 387-407.

Borland, R. (2011). Radical plumbers and PlayPumps: Objects in development. PhD.

Bratton, M. (2014). "Global TIES: Ten Years of Engineering for Humanity." International Journal for Service Learning in Engineering, Humanitarian Engineering and Social Entrepreneurship: 205-221.

Carberry, A. R., H. S. Lee and M. W. Ohland (2010). "Measuring engineering design self-efficacy." Journal of Engineering Education **99**(1): 71-79.

Chi, M. T. H. (2006). Two Approaches to the Study of Experts' Characteristics. The Cambridge Handbook of Expertise and Expert Performance. K. A. Ericsson, N. Charness, P. J. Feltovich and R. R. Hoffman, Cambridge University Press**:** 21-30.

Cooke, B. and U. Kothari, Eds. (2001). Participation: The new tyranny? New York, Zed books.

Coyle, E. J., L. H. Jamieson and W. C. Oakes (2005). "EPICS: Engineering projects in community service." International Journal of Engineering Education **21**(1): 139-150.

Design Council. (2019). "The Design Process: What is the Double Diamond?" Retrieved May 15, 2019, from https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond.

Drain, A., A. Shekar and N. Grigg (2017). "'Involve me and I'll understand': creative capacity building for participatory design with rural Cambodian farmers." CoDesign: 1-18.

Druin, A. (2002). "The role of children in the design of new technology." Behaviour & Information Technology **21**(1): 1-25.

Engineers Australia (2018). Entry into university engineering courses: 2018 update report, Engineers Australia.

Engineers without Borders Australia. (2018). "Community participation." EWB Knowledge Hacks Retrieved May 15, 2019, from https://ewbchallenge.org/community-participation.

Engineers without Borders Australia. (2019). "Implementing a Human Centered Approach."   Retrieved May 15, 2019, from https://www.ewb.org.au/blog/implementing-a-human-centered-approach.

Hazeltine, B. and C. Bull (1999). Appropriate Technology; Tools, Choices, and Implications. San Diego, CA, Academic Press, Inc.

Hess, J. L., J. Beever, A. Iliadis, L. G. Kisselburgh, C. B. Zoltowski, M. J. Krane and A. O. Brightman (2014). An ethics transfer case assessment tool for measuring ethical reasoning abilities of engineering students using reflexive principlism approach. 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, IEEE.

Hoffman, B. J., C. L. Kennedy, A. C. LoPilato, E. L. Monahan and C. E. Lance (2015). "A review of the content, criterion-related, and construct-related validity of assessment center exercises." Journal of Applied Psychology **100**(4): 1143.

Ilgen, J. S., I. W. Ma, R. Hatala and D. A. Cook (2015). "A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment." Med Educ **49**(2): 161-173.

Jesiek, B. K., S. E. Woo, S. Parringon and C. Porter (under review). "Development and initial validation of a situational judgment test (SJT) for global engineering competency (GEC)." Journal of Engineering Education.

Jolly, L., C. Crosthwaite, L. Brodie, L. Kavanagh and L. Buys (2011). The impact of curriculum content in fostering inclusive engineering: data from a national evaluation of the use of EWB projects in first year engineering. Australasian Association for Engineering Education Conference 2011: Developing engineers for social justice:

Community involvement, ethics & sustainability, Fremantle, Western Australia, Engineers Australia.

Kanji, N. and L. Greenwood (2001). Participatory approaches to research and development in IIED: Learning from experience. London, IIED.

Kilgore, D., C. J. Atman, K. Yasuhara, T. J. Barker and A. Morozov (2007). "Considering context: A study of first-year engineering students." Journal of Engineering Education **96**(4): 321-334.

Lucena, J. and J. Schneider (2008). "Engineers, development, and engineering education: From national to sustainable community development." European Journal of Engineering Education **33**(3): 247-257.

Mattson, C. A. and A. E. Wood (2014). "Nine principles for design for the developing world as derived from the engineering literature." Journal of Mechanical Design **136**(12): 121403.

Mazzurco, A. (2016). Methods to facilitate community participation in humanitarian engineering projects: Laying the foundation for a learning platform. PhD, Purdue University.

Mazzurco, A. and S. Daniel (in press). "Socio-technical thinking of students and practitioners in the context of humanitarian engineering." Journal of Engineering Education.

Mazzurco, A., S. Daniel and J. Smith (2019). Development of Socio-Technical and Co-Design Expertise in Engineering Students. Research in Engineering Education Symposium. B. Kloot. Cape Town, South Africa.

Mazzurco, A., J. L. Huff and B. K. Jesiek (2014). "The Energy Conversion Playground (ECP) Design Task: Assessing how Students Think About Technical and Non-Technical Considerations in Sustainable Community Development." International Journal for Service Learning in Engineering **9**(2).

Mazzurco, A. and B. K. Jesiek (2014). Learning from failure: Developing a typology to enhance global service-learning engineering projects. Proceedings of the 2014 American Society of Engineering Education Annual Conference and Exposition.

Mazzurco, A. and B. K. Jesiek (2017). "Five Guiding Principles to Enhance Community Participation in Humanitarian Engineering Projects." Journal of Humanitarian Engineering.

Mazzurco, A., J. A. Leydens and B. K. Jesiek (2018). "Passive, Consultative, and Coconstructive Methods: A Framework to Facilitate Community Participation in Design for Development." Journal of Mechanical Design **140**(12): 121401.

McKenna, A. F. (2007). "An investigation of adaptive expertise and transfer of design process knowledge." Journal of Mechanical Design **129**(7): 730-734.

McKenna, A. F., M. M. Hynes, A. M. Johnson and A. R. Carberry (2016). "The use of engineering design scenarios to assess student knowledge of global, societal, economic, and environmental contexts." European Journal of Engineering Education **41**(4): 411-425.

McMartin, F., A. McKenna and K. Youssefi (2000). "Scenario assignments as assessment tools for undergraduate engineering education." IEEE Transactions on Education **43**(2): 111-119.

Mitcham, C. and D. R. Muñoz (2009). The humanitarian context. Engineering in context. S. H. Christensen, B. Delahousse and M. Meganck. Aarhus, Denmark, Academica**:** 183-195.

Moskal, B. M., C. Skokan, D. Munoz and J. Gosink (2008). "Humanitarian engineering: Global impacts and sustainability of a curricular effort." International Journal of Engineering Education **24**(1): 162-174.

Murcott, S. (2007). "Co-evolutionary design for development: influences shaping engineering design and implementation in Nepal and the global village." Journal of International Development **19**(1): 123-144.

Pandian, S. R. (2004). A human power conversion system based on children's play. Technology and Society, 2004. ISTAS'04. International Symposium on, IEEE.

Passino, K. M. (2009). "Educating the Humanitarian Engineer." Science and engineering ethics **15**(4): 577-600.

Peeters, M.-C., E. Londers and W. Van der Hoeven (2014). "Design of an integrated team project as bachelor thesis in bioscience engineering." European Journal of Engineering Education **39**(6): 636-647.

Peng, K., R. E. Nisbett and N. Y. Wong (1997). "Validity problems comparing values across cultures and possible solutions." Psychological methods **2**(4): 329.

Russell, A. L. and L. Vinsel (2019). Make maintainers: Engineering education and an ethics of care. Does America need more innovators? M. Wisnioski, E. S. Hintz and M. S. Kleine. Cambridge, MA, The MIT Press.

Sanders, E. B. N. and P. J. Stappers (2014). "Probes, toolkits and prototypes: three approaches to making in codesigning." CoDesign **10**(1): 5-14.

Schneider, J., J. A. Leydens and J. Lucena (2008). "Where is 'Community'?: Engineering education and sustainable community development." European Journal of Engineering Education **33**(3): 307-319.

Shuman, L. J., M. Besterfield-Sacre and J. McGourty (2005). "The ABET "professional skills"—Can they be taught? Can they be assessed?" Journal of engineering education **94**(1): 41-55.

Sianipar, C., G. Yudoko, K. Dowaki and A. Adhiutama (2013). "Design Methodology for Appropriate Technology: Engineering as if People Mattered." Sustainability **5**(8): 3382-3425.

Smith, J., A. Mazzurco and P. Compston (2018). "Student engagement with a humanitarian engineering pathway." Australasian Journal of Engineering Education **23**(1): 40-50.

Ssozi-Mugarura, F., E. Blake and U. Rivett (2017). "Codesigning with communities to support rural water management in Uganda." CoDesign **13**(2): 110-126.

Thoma, S., W. P. Derryberry and H. M. Crowson (2013). "Describing and testing an intermediate concept measure of adolescent moral thinking." European Journal of Developmental Psychology **10**(2): 239-252.

Turner, J., N. Brown and J. Smith (2015). Humanitarian Engineering-What does it all mean? AAEE 2015: Blended Design and Project Based Learning: a future for engineering education, Deakin University.

Walther, J., N. W. Sochacka and N. N. Kellam (2013). "Quality in Interpretive Engineering Education Research: Reflections on an Example Study." Journal of Engineering Education **102**(4): 626-659.

Wood, A. E. and C. A. Mattson (2016). "Design for the developing world: Common pitfalls and how to avoid them." Journal of Mechanical Design **138**(3): 031101.

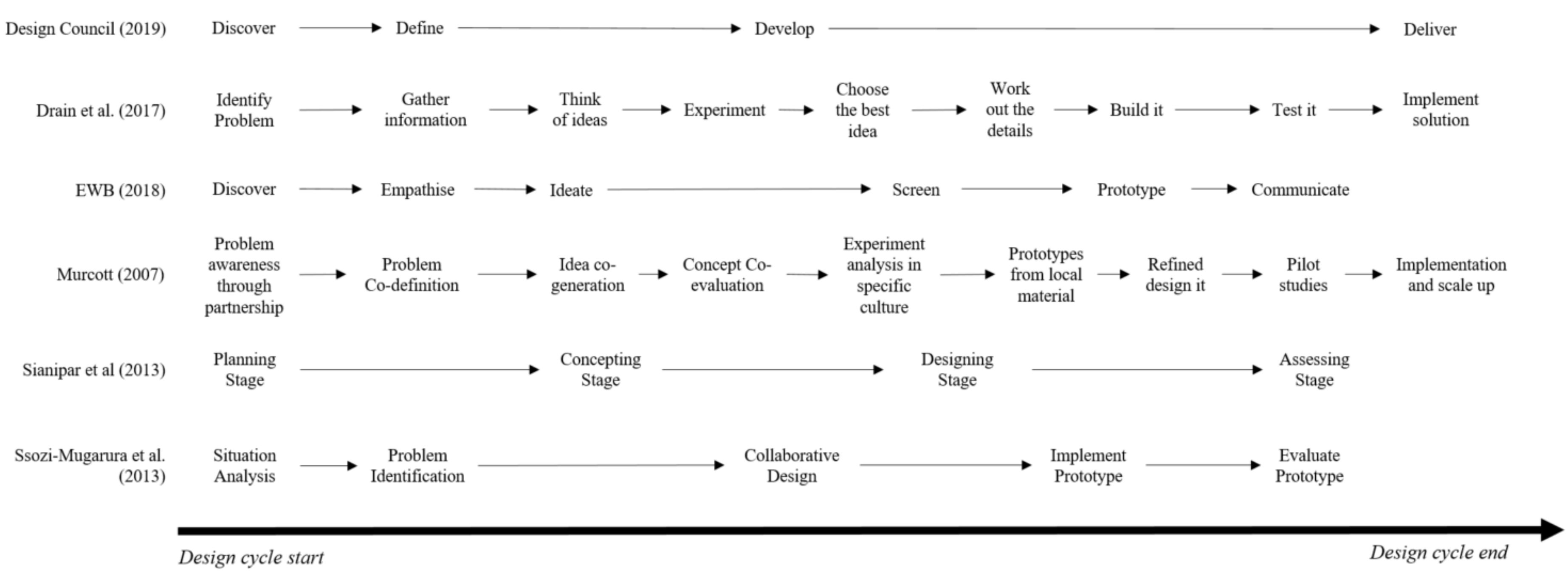

Figure 1. Different representations of the design process (note that while some of the original processes were modelled cyclically, here they are depicted linearly to facilitate comparison)

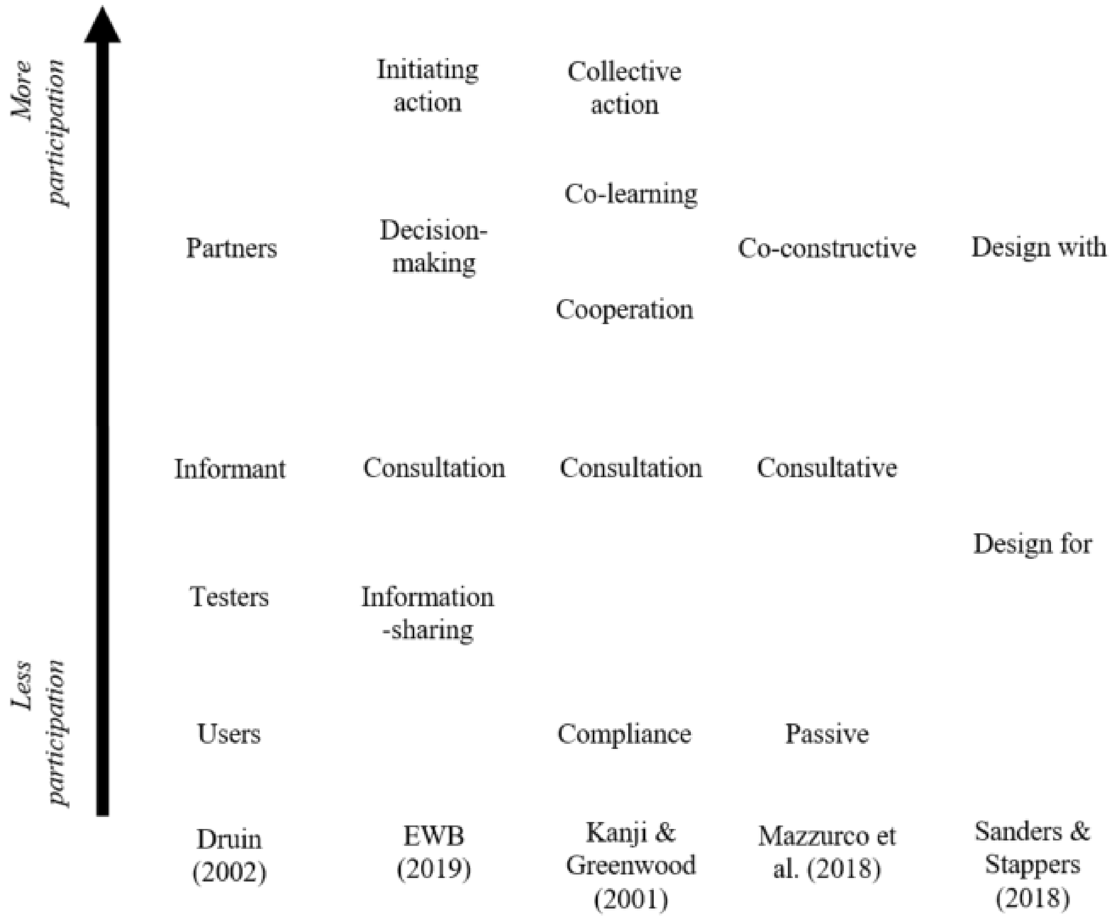

Figure 2. Different representations of community participation in humanitarian engineering design projects

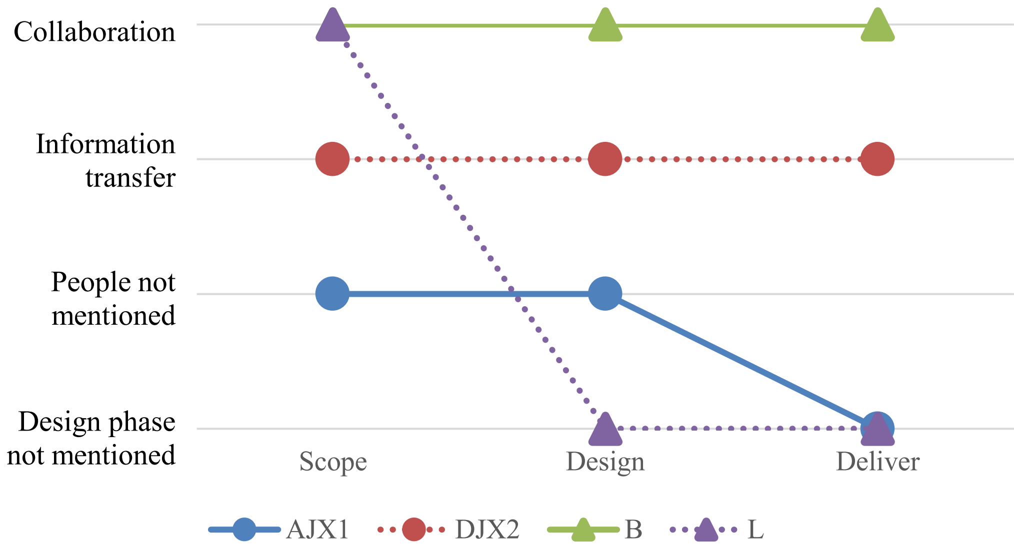

Figure 3. Comparing individual variation in co-design expertise (triangles refer to students, circles to practitioners)

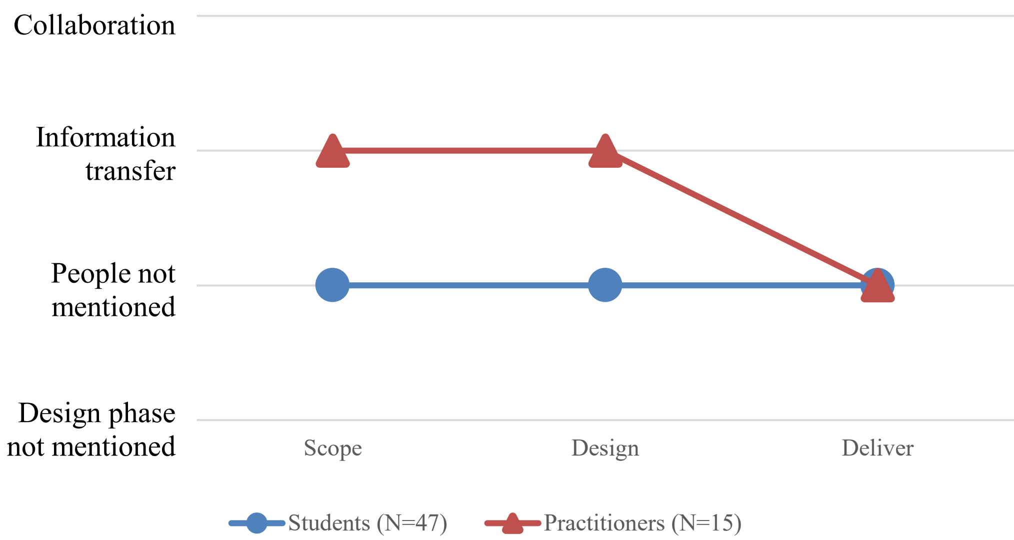

Figure 4. Median co-design scores (triangles refer to students, circles to practitioners)

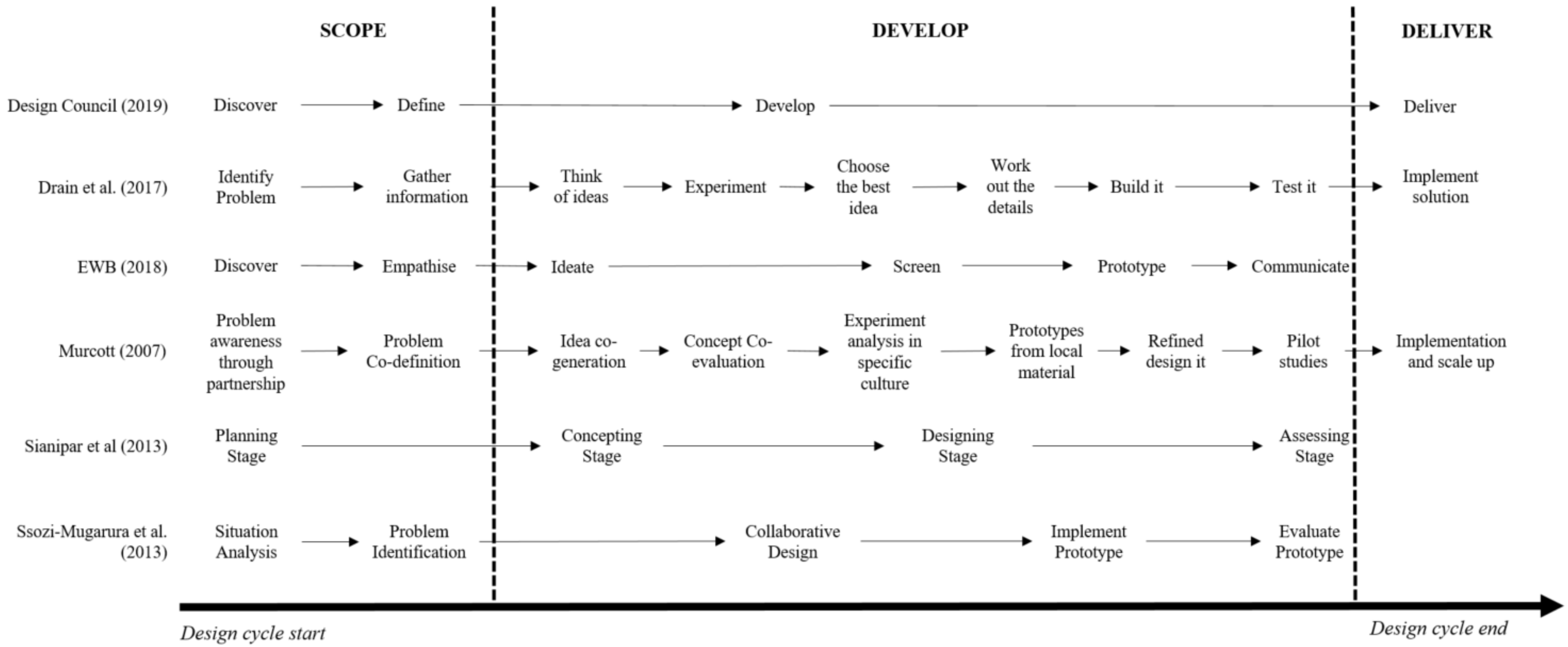

Figure 5. Evaluating our characterisation of design against the literature

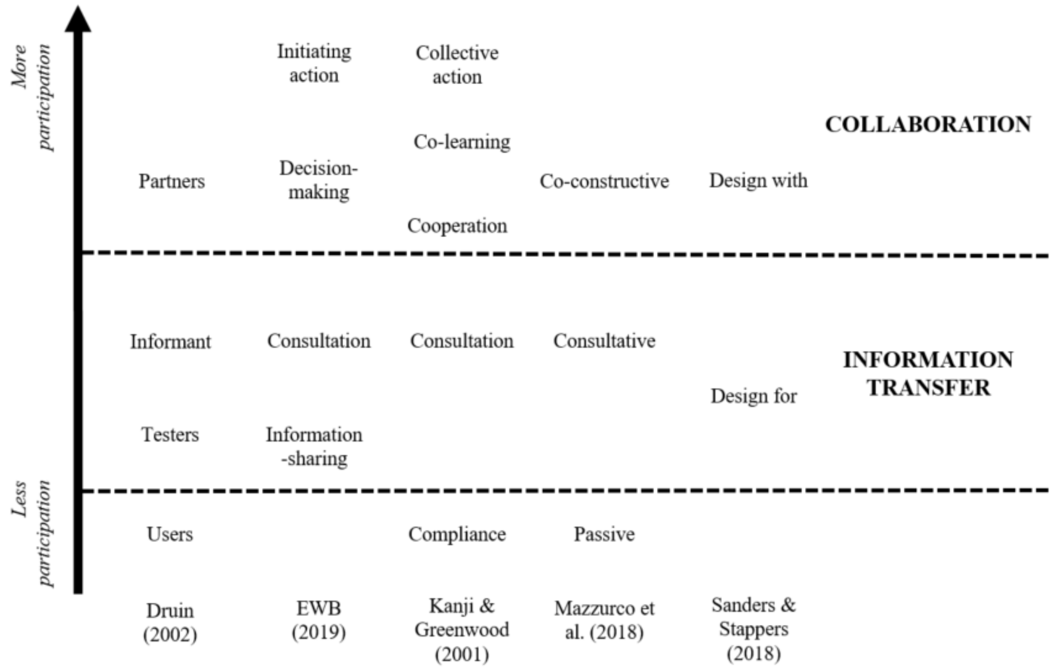

Figure 6. Evaluating our characterisation of community participation against the literature

Table 1: Practitioners' demographics and background

	N	*%*
Gender		
Male	9	56.3
Female	7	43.7
Non-binary	0	0
Highest degree		
High school	0	0
Bachelor	0	0
Masters	8	50
PhD	8	50
Subject studied at university		
Engineering	15	94
Other STEM subjects	4	25
Education	3	19
Arts	2	13
International development	2	13
Other fields	6	38
Context of HE experiences		
Taught HE at university	10	63
Conducted research on HE	13	81
Volunteered for HE-related organisations	13	81
Worked for HE-related organisations	7	44
Participated in international HE projects	11	69
Received training on HE	11	69
Time living or working overseas		
None	0	0
Less than 6 months	3	19
6 months to 1 year	1	6
1 to 2 years	3	19
3 years or more	9	56

Table 2. Percentage of agreement, correlation, and gradient of line of best fit between responses

	% agree	correlation	gradient
Rater 1 vs Rater 2	53%	0.64	0.53
Rater 1 vs authors' consensus	67%	0.74	0.72
Rater 2 vs authors' consensus	53%	0.73	0.84
Authors' consensus vs raters' consensus	60%	0.78	0.78

Table 3: Scoring rubric – design phases by community involvement.

<table>
<tr><th></th><th>Scope</th><th>Design</th><th>Delivery</th></tr>
<tr><td>Typical design activities</td><td>- Stakeholder research
- Context research
- Identifying design requirements</td><td>- Brainstorm / ideate
- Selection
- Prototype
- Test
- Iterate</td><td>- Produce, manufacture, build, installation
- Implementation
- Education
- Monitoring, Evaluation, and maintenance</td></tr>
<tr><td>Not mentioned (0)</td><td colspan="3">Did not include any reference to the activities typically belonging to the design phase.</td></tr>
<tr><td>Mentioned (1)</td><td colspan="3">Mentions at least one activity belonging to the design phases, but does not mention any form of engagement with stakeholders.</td></tr>
<tr><td>Information transfer (2)</td><td colspan="3">Mentions at least one activity belonging to the design phases and includes reference to some form of information transfer between engineers and stakeholders.</td></tr>
<tr><td>Collaboration (3)</td><td colspan="3">Mentions at least one activity belonging to the design phases and includes reference to a collaborative approach to doing so.</td></tr>
</table>

Table 4: Exemplar excerpts across the different design phases and levels of community participation.

Score	Scoping	Develop	Deliver
Mentioned (1)	Assess how much energy the school will need [AJX1]	Design a practical playground that works well [AJX1]	consider long-term maintenance issues [M]
Information transfer (2)	-get numbers of the operational needs -interview the school employees to get the typical work day schedule and needs -visit the site to see what kind of structures are possible to be built [I]	come up with some initial designs, compare those design based on how they perform in terms of the operational needs (the customer requirements that get translated into engineering characteristics). Get feedback from the customer [I]	Implement and capacity building. [G]
Co-constructive (3)	I would have a series of meetings/workshops with the community before even presenting this idea to see what their expectations/hopes/desires are around power for their school. I would want these workshops to happen in ways that could include a diverse cross-section of the community so that multiple voices are heard. I would want to talk through local resources (social and physical) that could be leveraged toward developing and maintaining a solution that fit the community's needs and desires. [L]	Collaborate with the local NGO and community to come up with an appropriate solution [A]	The detailed design will be released to tender to the local community. The contract that will engaged will be a build-operate-maintain contract that will allow the local contractor to put forth a proposal for the construction, commissioning, operation and maintenance of the infrastructure this will include expected budgetary spending. Where local contractors cannot be sourced consideration will be made to international contractors operating under an aid based or NGO framework that will work in partnership with the local community to engage and train local personnel for construction, operation and maintenance as much as possible. The contractor will be awarded and publically notified. The identified owner of infrastructure (whether that be local government or community group) will be tasked with raising the capital expenditure required to allow for the project construction and engaging the contractor under a legal contractor [O]

Table 5: Breakdown of pre-instruction students and practitioner co-design scores.

	Pre-instruction students (n=26)			Practitioners (n=15)		
	Scope	Design	Deliver	Scope	Design	Deliver
% of 0s	4%	8%	27%	13%	20%	27%
% of 1s	65%	58%	27%	20%	27%	40%
% of 2s	31%	35%	46%	40%	13%	20%
% of 3s	0%	0%	0%	27%	40%	13%