Design and Technology Education: An International Journal

29.3

Defining and Evaluating Argumentation Quality in the Context of Design Thinking: Using High School Students' Design Critiques from Foundational Engineering Courses

Wonki Lee, Purdue University, USA
Nathan Mentzer, Purdue University, USA
Andrew Jackson, University of Georgia, USA
Scott Bartholomew, Brigham Young University, USA
Amiah Clevenger, Purdue University, USA

Abstract

This research investigates students' argumentation quality in engineering design thinking. We implemented Learning by Evaluating (LbE) using Adaptive Comparative Judgment (ACJ), where students assess pairs of items to determine the superior one. In ACJ, students provided rationales for their critiques, explaining their selections. Fifteen students participated in an LbE exercise before starting their backpack design projects, critically evaluating multiple backpack designs and producing 145 comments. Writing comments required students to discern and justify the superior design, fostering informed judgment and articulation of their reasoning. The study used the Claim, Evidence, and Reasoning (CER) framework, adapted for engineering design thinking, to analyse these critiques. The framework emphasized three aspects: Empathy (understanding user needs), Ideation (deriving design inspiration), and Insight (gaining valuable understanding from evaluated designs). We employed both deductive and inductive content analysis to evaluate the argumentation quality in students' critiques. High-quality argumentation was identified based on six codes: user-focused empathy, design inspirations, logical rationalizations, multi-criteria evaluations, aesthetic considerations, and cultural awareness. Poor-quality argumentation lacked these elements and was characterized by vagueness, uncertainty, brevity, inappropriateness, irrelevance, gender bias, and cultural stereotyping. By identifying critical elements of effective argumentation and common challenges students may face, this study aims to enhance argumentation skills in engineering design thinking at the secondary education level. These insights are intended to help educators prepare students for insightful and successful argumentation in engineering design projects.

Keywords

LbE (Learning by Evaluating), ACJ (Adaptive Comparative Judgment), CER (Claim, Evidence, and Reasoning) framework, Design thinking

Introduction

There is an urgent need to enhance young individuals' comprehension of argumentation within a scientific context (Osborne et al., 2004). Osborne and colleagues (2004) asserted that science, a field that prides itself on rationality, often falls short in teaching students about the epistemological foundations of belief (Driver et al., 2000; Duschl & Osborne, 2002; Osborne et al., 2003). Oftentimes, this failure results in students' holding simplistic perceptions of science (Driver et al., 1996, 2000). Thus, one of the vital responsibilities for the current education is to cultivate argument construction and explanation. Such instruction enhances students' ability to comprehend and employ scientifically sound argumentation (Duschl & Osborne, 2002). Osborne (2004) underscored the significance of embedding scientific argumentation within classroom settings, suggesting it serves as a heuristic methodology that augments students' ability to navigate conceptual and epistemic objectives. This integration not only illuminates students' scientific cognition and reasoning but also facilitates formative assessment opportunities for educators. Consequently, the pursuit of epistemic objectives, such as the formulation, assessment, and refinement of scientific arguments, is imperative to modern science education.

Numerous studies have highlighted that fostering students' ability to construct scientific arguments can be significantly enhanced by presenting them with contrasting theories or evidence (Keogh & Naylor, 2000; Osborne et al., 2004; Settlage & Sabik, 1997; Solomon et al., 1992). This method not only strengthens critical thinking skills but also empowers students to integrate their scientific knowledge into discussions and decision-making processes. Importantly, the skills developed through constructing scientific arguments are not limited to science but are equally valuable in fields like engineering and technology design, where reasoned argumentation is critical for evaluating, justifying, and refining solutions (Dow et al., 2009; Erduran & Jiménez-Aleixandre, 2007; Jonassen, 2011; Osborne et al., 2004).

In the engineering education context, the introduction of competing evidence presented in parallel - rather than sequentially - has been shown to positively impact learning outcomes (Dow et al., 2009). For example, Dow et al. (2009) observed that parallel presentation fosters more diverse prototyping, allowing students to explore a broader range of design possibilities, ultimately improving the effectiveness of the design process. Similarly, Karabiyik et al. (2023) found that exposing students to contrasting cases activates prior knowledge, helping them identify and emphasize key elements of domain-specific concepts. This approach aligns with the work of Schwartz & Bransford (1998), who argued that contrasting scenarios encourage learners to focus on critical differences, thereby deepening their understanding of the material.

The Learning by Evaluating (LbE) method (Jackson et al., 2023; Mentzer et al., 2023) extends the concept of contrasting case analysis into engineering design thinking. This approach actively engages students in the evaluation of contrasting engineering designs items, which are presented in parallel. Then, they are asked to provide their rationale for selecting one over another. Students apply their domain knowledge to discern and judge contrasting design cases.

This immersive evaluative process not only involves making informed judgments but also requires students to articulate their reasoning, thereby deepening their engagement and understanding. By integrating such critical comparative analysis into the learning process, LbE echoes the positive influence of contrasting evidence on learning outcomes observed in scientific reasoning.

Additionally, through LbE, students are encouraged to explore a broader spectrum of design possibilities, enhancing their design thinking capabilities and the overall efficacy of the design process (Jackson et al., 2023; Mentzer et al., 2023; Thorne et al., 2024). LbE provides insights into design thinking at the onset of a project, rather than postponing experience of critiquing until students have delved deeper into their tasks. Importantly, in LbE, the examples for evaluation were instructor-curated to align with specific learning goals or project components. Based on this strategy, researchers posit that students will not only gain a clearer direction for their project, but the comparative evaluations they undertake will also enhance their decision-making abilities as well as argumentation skills, especially when faced with open-ended challenges.

In an engineering design context, Strimel et al. (2021) explored the application of LbE to shape engineering students' design choices. A key component of this process is comparative judgment (CJ), which involves systematically comparing pairs of student work based on defined criteria to determine which is of higher quality. Unlike traditional grading, CJ relies on expert or peer evaluations through a series of binary comparisons, which are aggregated to provide a rank order of work. This approach not only reduces subjective bias but also provides valuable insights into the relative strengths and weaknesses of designs. The research findings suggest that the CJ procedure within LbE allowed students to gain valuable perspectives for improving their designs by evaluating the work of their peers, although direct feedback on their own projects was not provided as part of the process.

Mentzer et al., (2023) recently examined the nature of students' reasoning and comments using computer-assisted content analysis, complemented by a subsequent qualitative content analysis. Their research uncovered a spectrum of critical and scientific thinking skills among high school students in the realm of engineering design thinking. The CER framework, which emphasizes Claim, Evidence, and Reasoning as foundational components of constructing scientific arguments, provides a structured approach to analyzing student argumentation. However, the study applied the CER framework in its original form, without customizing or expanding it specifically for the engineering design thinking context.

Given that the CER framework is domain-specific, this study aims to identify and assess argumentation quality specifically within the engineering design context. This is explored as students engage in LbE and articulate their critiques in the comment section (see Figure 1).

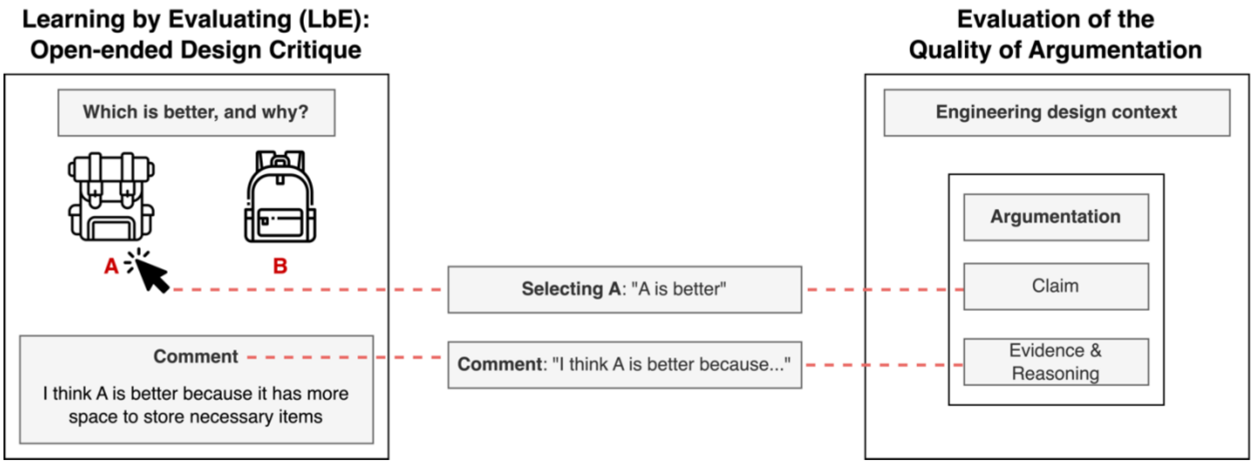

Figure 1. Research design: LbE and the evaluation of the quality of argumentation

Since CER framework varies depending on the domain (Slavit et al., 2021), this study aims to delineate and assess the quality of argumentation within the context of engineering design, particularly when students engage in the LbE and articulate their critiques in comments. This investigation will extend beyond merely evaluating the presence and elaboration of claims, evidence, and reasoning. It will also explore how these argumentation elements are applied within the essential criteria of the design thinking process. The research question for the present study is outlined as follows.

Research Question:

> *How is high-quality argumentation defined and operationalized within the context of engineering design thinking, and conversely, what characterizes poor-quality argumentation in this field?*

Examining high-quality argumentation within the context of design thinking can serve as an exemplary model for students, guiding their analytical development. Conversely, identifying instances of weak or poor-quality argumentation can spotlight specific areas that require enhanced instructional focus. By reviewing students' comments in conjunction with the categorizations of quality presented in this article, teachers can gain valuable insights into their students' understanding and reasoning. This dual approach not only helps in recognizing the current level of students' argumentative skills but also aids in determining where additional instruction or support may be necessary to foster improvement.

Literature Review

Adaptive Comparative Judgment (ACJ)

Adaptive Comparative Judgment (ACJ) is a technique developed by researchers and psychologists to evaluate complex or subjective tasks, such as open-ended responses, design projects, or creative work, through a series of pairwise comparisons. It builds upon the

Comparative Judgment (CJ) approach developed by Thurstone in 1927. Thurstone's framework posits that judges can make more precise discriminations between two items when comparing them directly, rather than assigning absolute quality scores, which can be influenced by individual biases or environmental factors. In both CJ and ACJ, judgments are made independently of others and rely on the relative quality of each pair rather than pre-defined scoring rubrics (Bartholomew, 2021). By focusing on comparative evaluations, ACJ offers a reliable and efficient method for assessing tasks that are difficult to quantify using traditional scoring approaches.

In the transition to the twenty-first century, Pollitt utilized comparative judgment, integrated multifaceted statistical analyses, and introduced the approach as Adaptive Comparative Judgment or ACJ. Literature demonstrates greater validity, reliability, and robustness of ACJ compared to CJ (Kimbell, 2022; Mentzer et al., 2021; Pollitt, 2015)-. This approach is considered "adaptive" as it employs an algorithm which assigns values to each of the objects being compared, later showing the judge comparisons of similarly valued items to fine tune the rank order rather than comparisons of paired items that are already determined to be extremely higher or lower than each other.

As an example, in 2012, Pollitt conducted a study involving 1,000 writing samples from students aged nine to eleven (Pollitt, 2012b). In this research, educators were recruited to act as judges and underwent comprehensive training. Subsequently, they were assigned the task of evaluating pairs of writing samples and determining the superior one from each pair. As evaluators engage in judgments, examples are associated with a value of one or zero (winner or loser), and then put through another series of comparisons based on similar values after the initial random judgments of pairs were made. This continued until all writing pieces were given a final rank, continually compiling based on their previous results. At the end of the test, all judges that gave feedback reported that they would rather use ACJ than their traditional form of marking (Pollitt, 2012b).

With the promise of ACJ firmly in view, researchers are actively exploring avenues for its broader application in technology, engineering, as well as design and technology fields throughout the United States (Bartholomew, 2021). Within higher education, ACJ has garnered significant attention, particularly in the context of design-related work (e.g., portfolios, presentations, and prototypes). In the context of design learning, Kimbell & Stables (2007) have explored the broader implications of comparative judgment, emphasizing its potential to enhance design-based education by fostering critical thinking and reflective practices. Their research has provided foundational insights into how comparative methodologies, such as ACJ, can be integrated into engineering and design education to promote robust assessment practices. Also, Bartholomew & Strimel (2018) examined the application of ACJ in engineering education to assess students' design portfolios. Their study highlighted how ACJ not only streamlined the evaluation process but also provided robust reliability and validity in ranking the quality of design solutions, enabling educators to identify and reward nuanced differences in creativity, functionality, and innovation. Beyond design, the potential of ACJ extends to other disciplines, including mathematics, audio content, and graphic design, demonstrating its versatility in evaluating complex, subjective tasks (Bartholomew, 2021).

Learning by Evaluating (LbE)

Learning by Evaluating (LbE) is an educational approach based on ACJ. Previous studies that incorporated students as judges in the ACJ process (e.g., Bartholomew, 2019; Bartholomew, 2018;) recognized the contribution to learning of this participation; additionally, studies wherein critique in ACJ was made *during* a project, as opposed to the end of the project, have enabled students to apply insights from the comparative process (Bartholomew et al., 2019). Therefore, LbE amplifies these paradigmatic changes to the traditional educational experience by using *student judges in the beginning* of the design process to prime student learning. In the realm of academic assessment, LbE introduces a novel approach that empowers students to take a proactive role in their learning process. It serves as a valuable tool for them to enhance the quality of their later work, stimulate creativity in design, and effectively leverage feedback (Bartholomew et al., 2019).

Our team, supported by a National Science Foundation (NSF) grant—a U.S. program funding research and education in science and engineering—is investigating the optimal implementation of LbE to achieve desired student learning outcomes. For example, in the study conducted by Bartholomew et al. (2020), the research group introduced LbE in an entry-level college class at the beginning of their task. In this case, students were to craft Point Of View (POV) statements which help lean into designing a future solution. Students in the experimental group examined POV statements from previous years, made side-by-side comparisons, and then composed their own statements. Subsequently, students assessed their peer's statements and those of the control group, assigning rankings. The results showed that the treatment group received higher rankings on average, with 7 of the top ten POV statements originating from this group.

In secondary education, Bartholomew's 2019 study focused on implementing LbE in a seventh-grade classroom at the middle and end of the project, hence allowing students to view feedback, make corrections of their designs, and be re-evaluated at the end. Once again, the treatment group demonstrated significantly higher rankings compared to their peers who had not participated in the experiment. Furthermore, the majority of students expressed satisfaction with the ACJ process, highlighting their enjoyment of learning from peers and engaging with constructive feedback (Bartholomew et al., 2019). Literatures thus consistently demonstrates the positive outcomes of LbE as a part of the learning experience for students.

Students' Learning Through Scientific Reasoning

Scientific reasoning is a problem-solving approach rooted in critical thinking, where individuals employ all available information to arrive at informed conclusions. Scientific reasoning is related to cognitive abilities such as critical thinking and reasoning which is used when developing critical STEM activities such as developing experiments, creating hypotheses, and deducting outcomes (Bao et al., 2009). Developing scientific reasoning will help students to solve future relevant problems in the STEM context. Further, scientific reasoning is a common practice shared in science and engineering, important for applying this to STEM education. It also has a long-term impact on students' academic achievement (Bao, 2009). A teacher can cultivate a student's scientific reasoning skills by having them participate in experimental method and free inquiry learning and modification inquiry (Khoirina, 2018)

Scientific reasoning is an approach to problem-solving that is deeply anchored in critical thinking (Stephens & Clement, 2010). It involves using all available information to draw informed conclusions. This form of reasoning is intricately linked to cognitive abilities like critical thinking and reasoning, which are crucial in developing essential STEM activities, such as designing experiments, formulating hypotheses, and deducing outcomes (Bao et al., 2009). It enhances the immediate learning process and has a lasting impact on students' academic achievements. Additionally, by fostering scientific reasoning, students can effectively tackle relevant future problems within the STEM context. Moreover, as a common practice in science and engineering, the integration of scientific reasoning into STEM education is vital.

Argumentation Quality Within the Context of Engineering Design Thinking

This research evaluates the quality of scientific argumentation using the Claim, Evidence, Reasoning (CER) framework proposed by McNeill and Krajcik (2008) which concentrates on three central aspects of an argument. This framework itself is an adaptation of Toulmin's argumentation pattern, which encompasses claims, rebuttals, and justifications (Toulmin, 2003). As defined by Wallon et al., (2018), the three integral components are (a) presenting a **claim** that addresses a specific question, (b) offering **evidence** that supports the claim, and (c) providing **reasoning** grounded in scientific principles to elucidate how the evidence substantiates the claim. These components are frequently recognized as fundamental aspects of robust, high-quality argumentation.

Slavit et al., (2021) expanded upon the concept of the CER framework, suggesting that the manner in which claims are reasoned can vary across disciplines. They posited that while the fundamental concepts and principles like scientific notions remain consistent, what constitutes a quality argumentation might differ significantly from one discipline to another. For example, they highlighted the distinct nuances in the principles of argumentation across mathematics, science, and engineering. (See Figure 2.) These differences align with paradigms for professional thinking across disciplines (Cross, 1982; Honey et al., 2014; Kelley & Knowles, 2016).

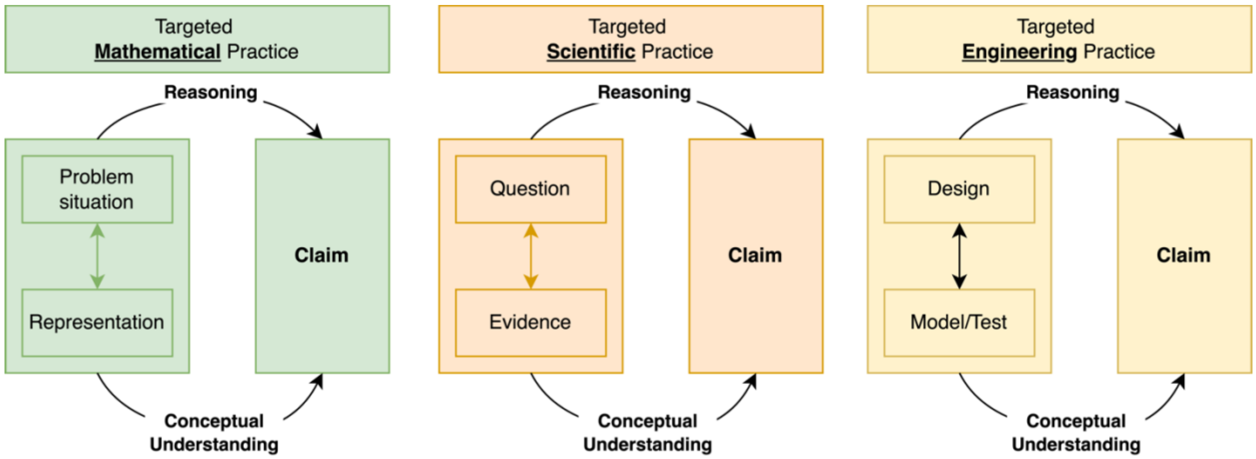

Figure 2. Disciplinary ways of thinking; using reasoning to make claims in mathematics, science, and engineering (Slavit et al., 2021)

In the LbE and engineering design context of this research, we redefined the components of scientific argumentation. Initially, a student's '*claim'* is interpreted from their preference for 'A' over 'B' or the converse, when making a comparative judgement. Consequently, their selection is conceptualized as the claim, 'Design/Artifact A is superior to B.' The image presented to the students is considered as a '*design'*. The '***model/test***' draws upon students' personal experiences and the conceptual knowledge they have acquired from their exposure to engineering design thinking thus far, because the LbE task occurs prior to the design project. Thus, students are asked to draw upon their personal experiences to conceptualize and assess the design's merits, visualizing its viability within their specified engineering design practices.

Beyond the basic components proposed by Slavit et al. (2021), evaluating argumentation quality within a given context permits the incorporation of conceptual features of the discipline. High-quality reasoning within argumentation should not only present claims, evidence, and reasoning but also demonstrate an understanding of the disciplinary practices that shape the argumentation. In the context of engineering, this includes the integration of the design-thinking process, which is central to the discipline. Students have received education in a design thinking process analogous to the methodology outlined by IDEO (2013), emphasizing its iterative and non-linear nature. Consequently, we have incorporated this process into the engineering argumentation framework as illustrated in the third image from the left in Figure 2. This integration supersedes the "design ↔ model/test" (see Figure 3). Therefore, the five stages of the design thinking process — empathize, define, ideate, prototype, and test (Dam, 2023) — have been employed to instruct students in engineering design. We posited that this non-linear and iterative procedure offers a more comprehensive representation of the engineering design thinking processes that students employ during their argumentation.

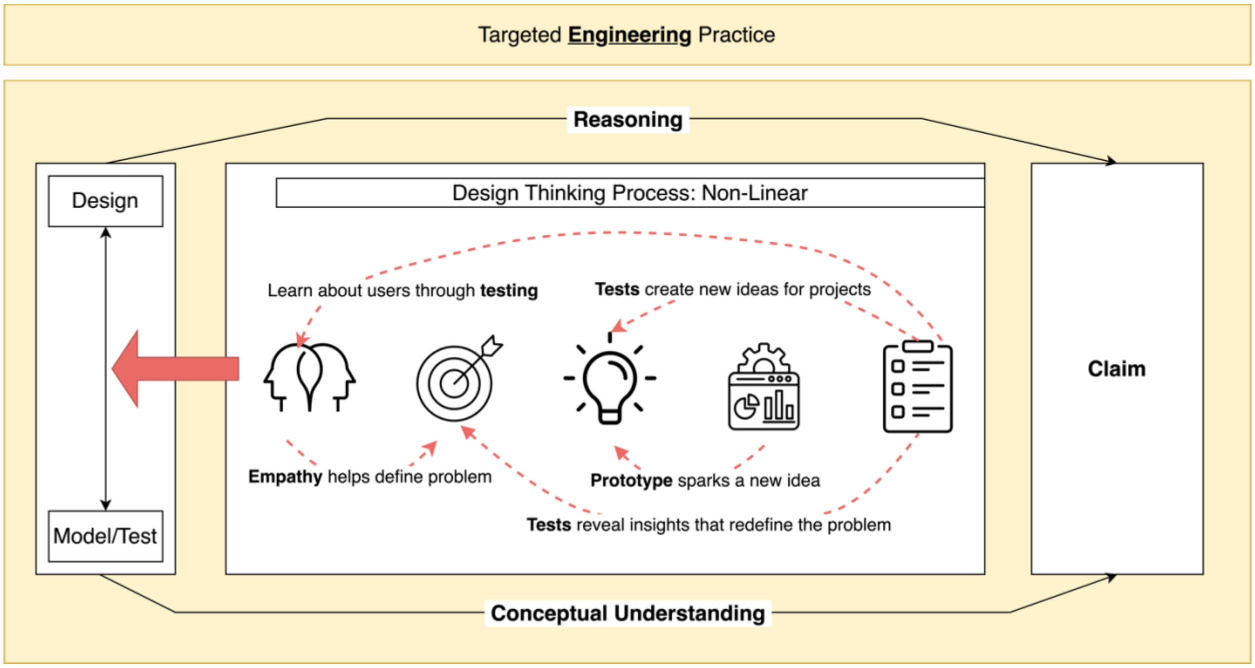

Figure 3. Engineering design thinking, embedded with a non-linear process

By integrating the iterative design thinking process into our reasoning approach, we were able to more precisely define the 'design ↔ model/test.' Subsequent team discussions led to a deeper understanding of how this process can be effectively applied in the current context and the distinct significance of each step within our framework. Figure 4 showcases the completed argumentation framework, devised for appraising the quality of students' reasoning pertinent to the present research. Initially, the framework was integrated within the discipline of engineering design. Subsequently, it was tailored to suit the specificities of the engineering design context, achieved through the application of an engineering design process. This process was instrumental in enhancing the 'design ↔ model/test' component, thereby incorporating an element of conceptual appropriateness. Consequently, we determined that high-quality argumentation is exemplified not merely by the presence of claims, reasoning, and evidence, but also by the inclusion of empathy, ideation, and insight within the students' reasoning process.

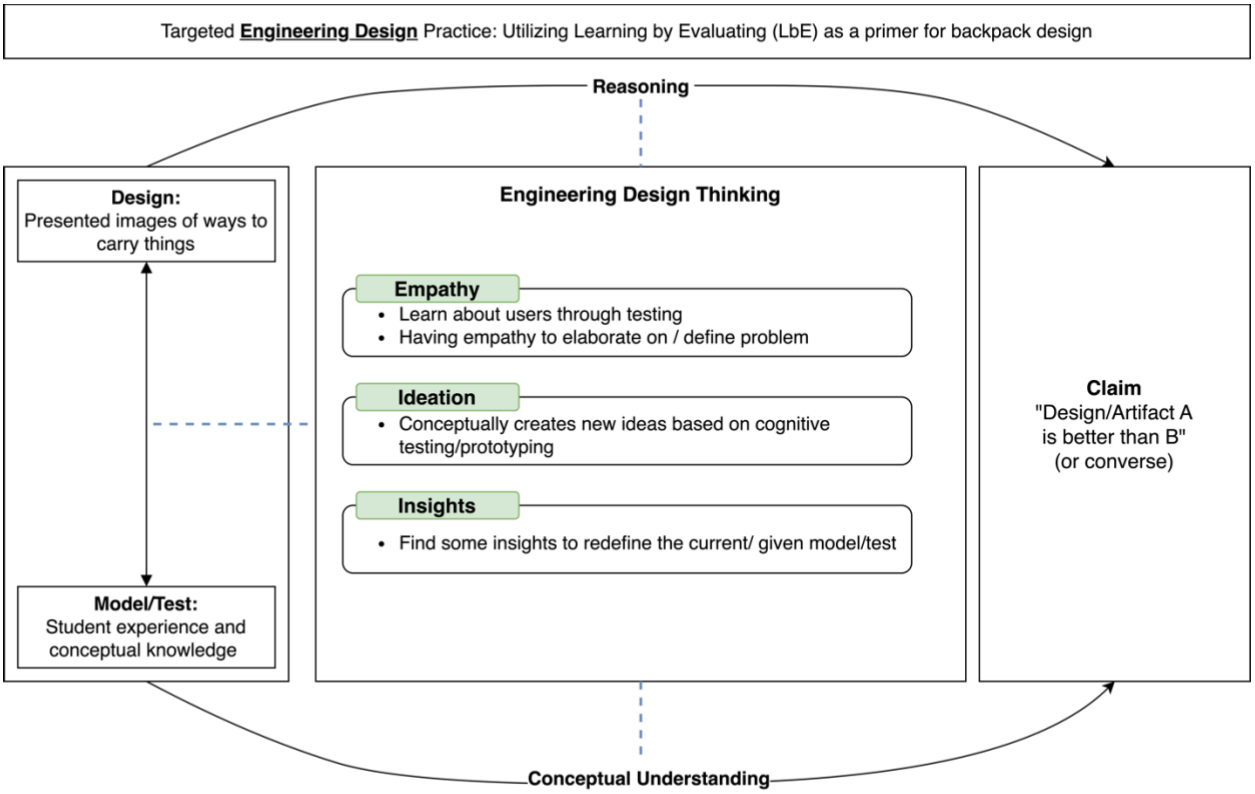

Figure 4. A framework for evaluating student argumentation quality within an engineering design thinking context

Methods

Research Contexts and Participants

The present research collaborates with DeKalb County Schools in Atlanta, a major urban school district. The district features a diverse student body, representing over 155 nationalities and exhibiting proficiency in more than 185 languages, as highlighted in the district's report(DeKalb County School District, 2021). Notably, 95% of the students participating in this research belong to groups that are historically underrepresented in STEM disciplines, predominantly Black and Hispanic populations (Artiles et al., 2005; Cain, 2012). For the Spring 2022 implementation of the project, five schools within the district participated. This included one teacher from each school (*N* = 5) and their respective students (*N* = 196), all of whom were enrolled in the foundational engineering design course, Foundations of Technology (FoT), authored by the International Technology and Engineering Educators Association's STEM Centre for Learning and Teaching.

Prior to the data collection, consent and assent forms were distributed to students, teachers, and parents/guardians, as required by the researchers' University Institutional Review Board and the school district's research coordinators. This study was implemented when teachers were delivering a similar challenge to their FoT classes—to complete a backpack design. A total of 15 students from three schools received full consent from their schools, teachers, and parents. These 15 students provided a total of 145 comments on the artifacts.

Measures and Procedures: Backpack Comparison and Critique

The overarching procedure for the present LbE session is delineated in Figure 5. The objective of this task was to encourage students to enhance their ideation process within the framework of design thinking. The web-based software, RMCompare, was utilized for this study. In the current study, the software facilitated comparative assessments, enabled commenting, and displayed the final rank during the debrief discussion.

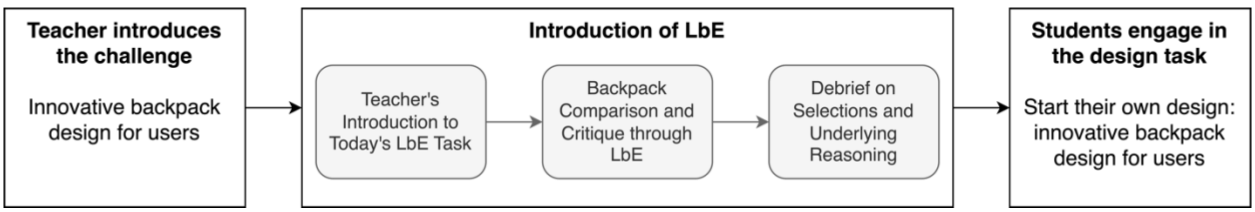

Figure 5. LbE session procedure: Backpack comparison and critique

Researchers and teachers uploaded a total of 39 diverse design items to the RMCompare software from various sources, all related to the objective of effective backpack designs. These items included images of actual backpacks as well as other visuals associated with backpacks that have the capacity to carry or hold items. The LbE task required students to select the example from a pair that they found most helpful or inspiring for their ideation process. This task provided students with an opportunity to evaluate and determine which option is superior and how it contributes to their ideation for a new product design. When introducing the LbE session, teachers prompted students to ponder the question 'What constitutes a superior backpack?' based on the following queries:

- What characteristics define an excellent backpack?
- How can each example serve as a source of inspiration for solving this problem?
- Which example might have transferable elements you want to use in your design?

Students then viewed pairs of images. In each case, they were required to select which they believed to be the better option (claim) to help them consider ways to improve their own backpack designs (see Figure 6). Right after their decision, students were required to justify their choice of one item over another. Each student made an average of 5 comparisons with a minimum of 3 and a maximum of 7 judgments. The entire experience spanned approximately 25 to 30 minutes, with the LbE comparison and critique constituting 10 to 15 minutes of that duration.

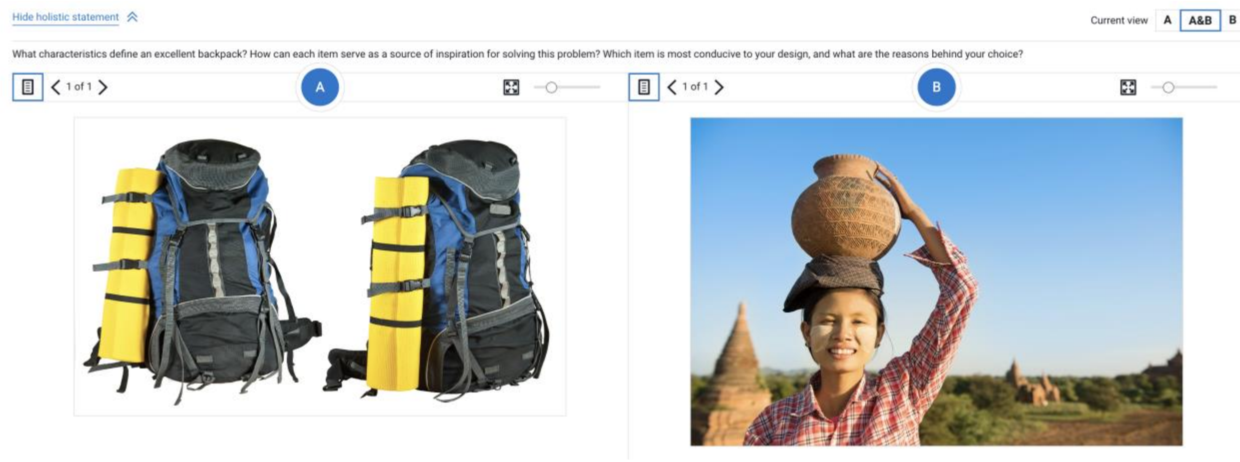

Figure 6. Screenshot of Backpack comparison session. "Left (A): ©Ivan/ Adobe Stock #337998470", "Right (B): ©cegli/ Adobe Stock #27278935"

At the end of the LbE task, teachers presented the final ranks in a wrap-up discussion. Students actively engaged in the conversation, discussing their judgments of the images, their reasons for selecting or not selecting them, and the rationale behind their choices.

Data Analysis

We employed qualitative content analysis as described by Krippendorff (2018) and further refined by Hsieh and Shannon (2005). The data analysis comprised two phases: deductive and inductive. Specifically, the data consisted of 145 comments provided by 15 students, which were analysed to identify patterns and themes related to their feedback on design items and their reasoning processes. This methodological approach, both integrative and iterative, draws from Andersson et al., (2015). It aligns with the theoretical assertion that content analysis's primary advantage lies in its adaptability to various research designs, tailored to specific research objectives (Elo & Kyngäs, 2008). Before analysing the data, we assessed whether the individual piece of feedback or reasoning provided by a student during the backpack design comparison session accurately represented the population. This involved revisiting the same backpack comparison session from the prior semester.

The Deductive Phase

We employed a deductive content analysis approach, following Krippendorff's (2018) guidelines. This method, structured on prior knowledge, suits our needs as we analyze argumentation within the engineering design thinking context. Before coding, researchers reviewed the theoretical framework for deductive content analysis (refer to Figure 4), as previously detailed. The broad categorization used as a framework in the analysis of the texts is as below (see

Table 1).

Table 1. Categorization matrix (i.e., codebook): Quality argumentation within the engineering design context

Quality argumentation criteria 1.	Quality argumentation criteria 2.	Categories based on criteria 1 and 2	Codes
Has claim, evidence, and reasoning	Appropriate for engineering design thinking context	Empathy	• Learn about users through testing • Having empathy to elaborate on or define the problem
		Ideation	• Conceptually creates new ideas based on cognitive testing/prototyping
		Insights	• Find some insights to redefine the current/ given model or test

The analysis began with in-depth reviews of written feedback transcripts retrieved from the RMCompare software, which captured students' comments and reasoning during the backpack design comparison sessions. We segmented the main coding phase into multiple rounds, coding smaller units between reliability and consensus checks.

Initially, researchers randomly selected and coded 10% of the comments using pre-defined categories from

Table 1, highlighting and coding relevant text segments. This was followed by a discussion to evaluate the framework's suitability. Subsequently, coding extended to an additional 30% of responses. The use of predetermined categories, a codebook, and a coding sheet aimed to minimize subjectivity in the coding process. After this second phase, researchers worked individually, categorizing the remaining codes to establish an initial classification. This approach revealed consistent patterns in the data, confirming the validity of the categories and their applicability across diverse responses. The findings also highlighted areas where further refinement of the categories might enhance their precision and utility for future analysis.

The Inductive Phase

Incorporating inductive analysis, researchers sought a deeper and more nuanced understanding, going beyond initial text categorizations. Specifically, this approach addressed two key limitations identified in the earlier categorization: 1) Identification of more specific sub-codes: During analysis, it became evident that more detailed sub-codes were necessary. This required the development of additional, more granular coding levels. 2) Exploring poor-quality argumentation: Existing categories failed to adequately explain instances of poor-quality argumentation. Consequently, there was a consensus to further delve into and elaborate on aspects of argumentation lacking in quality.

Units of analysis were abstracted into codes, considering their similarities and differences as outlined in

Table 1. Following collaborative discussions, these codes were categorized and regrouped according to content. This phase facilitated further abstraction, leading to the identification of three new subcategories under the ‘insights’ code. We also discovered a new code that shared characteristics of both empathy and insights. Furthermore, seven categories were newly established to specifically address instances of poor-quality reasoning.

Results

Researchers identified 13 distinct codes, classified into two categories based on the quality of argumentation: ‘Quality argumentation’ and ‘Poor-quality argumentation’ (see

Table 2). ‘Quality argumentation’ denoted desirable comments, which includes empathy, ideation, and insights. ‘Poor-quality argumentation’ designated comments which lack empathy, ideation, and insights. Besides these discipline-related aspects, ‘Poor-quality argumentation’ also referred to comments lacking clarity, essential components of reasoning, or displaying biases or stereotypes.

During the inductive phase, within the ‘Quality argumentation’ category, we identified six different subcodes under three main codes. Under empathy, we discovered one subcode: user-centred empathy. Under ideation, we identified one subcode: elucidation of design inspirations. For insights, three subcodes emerged: a) explanations for choices, b) detailed responses encompassing multiple criteria, and c) articulation of aesthetic attributes. Additionally, a code that encompassed both empathy and insights was identified: explanations of cultural awareness.

Conversely, in the Poor-quality argumentation category, seven distinct codes were pinpointed: 1) vagueness, 2) uncertainty or lack of knowledge, 3) excessively brief answers, 4) inappropriateness, 5) off-topic responses, 6) gender bias, and 7) cultural stereotyping. Notably, some codes exhibited overlap, especially in poor-quality argumentation. For instance, vague comments often coincided with being terse and short, complicating their interpretation.

Table 2. Overview of the content analysis: 'Quality argumentation' and 'Poor-quality argumentation'

Category	Code	Sub-code	Explanation	Example*	Occurence (%)
Quality argumentation	Empathy	User-Centred Perspectives	A comment from a design thinking perspective, which demonstrates a deep understanding and consideration of the user's perspective.	Choose B would be a good bag to bring with you for hiking and climbing up rough surfaces.	16 (10.94)
	Ideation	Explanation of design inspirations	A comment explaining inspirations from the options that have guided their engineering design process.	Helps show that my design should come with multiple parts.	7 (4.97)
	Insights	Reasoning for selecting and not selecting	A comment offers a direct comparison with explanations for why certain choices were made and why other options were not chosen.	B is more unusual and unique. However, A is more streamlined.	6 (4.4)
		Detailed comment with many criteria	A comment that provides a thorough and comprehensive analysis or evaluation by considering multiple factors or criteria.	Having a small and easy-to-carry book bag to carry around is pretty useful. Plus, the old backpacks can make your back or shoulder hurt.	7 (4.97)
		Explanation of aesthetic features	A comment of aesthetic features typically delves into the characteristics, qualities, or elements that contribute to the visual appeal, beauty, or artistic aspects of an engineering design.	I like this one better because it has more color and design.	5 (3.64)
	Empathy/ Insights	Explanation of cultural awareness	A comment provides insights into an individual's or group's understanding and recognition of various cultural aspects, practices, and values.	Option B is widely used in many traditional cultures, but seems less efficient and looks unhealthy compared to modern backpacks.	9 (6.18)

Poor-quality argumentation	Lacks empathy, insights, and ideation	Vague	A comment that lacks clarity, specificity, or detail, making it difficult to understand the intended meaning.	Cuz Mmm	23 (16.17)
		Uncertainty/ Lack of knowledge	Comments provided when a student is unsure or does not possess the necessary information or knowledge.	idk I have no idea Don't know	15 (10.43)
		Too Short Response	A comment is too brief or lacks sufficient detail.	much space don't close	15 (10.43)
		Inappropriate answer	A comment that is not suitable, expressing blaming or fault-finding.	Looks just dumb Bad design Wrong design	9 (6.18)
		Off-topic	Unrelated or not relevant to the subject or question.	My mom says I am so special! Ride	8 (5.31)
		Gender biased	A comment that exhibits prejudice or discrimination based on one's gender.	Who would use a girly laced bag as a way of carrying your stuff?	8 (5.31)
		Cultural stereotyping	A comment that perpetuates or promotes stereotypes about a particular culture or ethnic group.	Backpacks like B are smarter to use instead of commercially sold items made in country B that are incredibly low quality. Bags made in country A is ALWAYS good and fancy. They are always the best!	5 (3.64)

****Note***. Certain responses in the example may contain grammatical errors as they are direct quotations from students' responses.

Quality Argumentation

Quality argumentation in this context is defined by three key criteria (i.e., empathy, ideation, and insights), as outlined in our theoretical framework and consistent with engineering practices (see Figure 4). Empathy is highlighted when students demonstrate an understanding of user needs or apply empathy to improve a design. Ideation is valued for argumentation that shows how students can modify the presented design to enhance their own models. Insights are recognized when students extract useful information from a design, identifying its strengths and weaknesses. This framework of high-quality argumentation is further expanded upon in the subsequent content analysis. Here are detailed explanations of each sub-code, accompanied by quoted excerpts from student argumentations. Please note that some responses might contain grammatical errors, as they are direct quotations from students' responses. Additionally, to streamline the analysis, each category has been abbreviated (e.g., 'Empt' for 'Empathy') and sub-categorized with sequential numbering (e.g., 'Empt-1'). These labels are used throughout the analysis to reference specific examples.

Empathy (Empt): User-Centred Perspectives

- **Empt-1:** "…choose B would be a good bag to bring with you for hiking and climbing up rough surfaces…"
- **Empt -2:** "…the length of the user's hike is the first thing to consider. For a day hike, a pack between 10 and 25 liters should be enough. You could fit your water bottle and your picnic in this pack, as well as a jacket and sunscreen, to cater for all weather conditions…"
- **Empt -3:** "…a good bag to bring with you for hiking and climbing up rough surfaces…"
- **Empt -4:** "…arms can get tired of holding baby so you can hold him using your backpack and you also have space to put your things…"
- **Empt-5:** "…book bag becomes really heavy during the first day of school because of the number of books I carry…"

From a design thinking perspective, the 'user-centred perspectives' sub-code emphasizes empathy for users they posit for a design. It reflects a profound understanding and consideration of the user's needs. In this sub-code, students exhibited reasoning from various user viewpoints. The first three comments (Empt-1 to Empt -3) provided examples illustrate empathy towards hikers, showing an understanding of the challenges and environments they encounter while using the backpack. Besides of the hiker, in their design selection rationale, the argumentation in the user-centred perspectives sub-codes also tend to consider other user groups, such as parents with babies (Empt -4), and students carrying many books (Empt -5), demonstrating a broad application of empathy in their argumentation.

Ideation (Idtn): Explanation of design inspirations

- **Idtn-1**: "…helps me show that straps could be involved in my design…"
- **Idtn -2**: "…help the idea that my backpack design should be able to mount onto other objects…"
- **Idtn -3**: "…show that my design should be able to hold heavy items without breaking…"
- **Idtn -4**: "…shows that my design should come with multiple spaces…"

'Ideation: Explanation of design inspirations' encompasses students' comments that highlight the understanding derived from the options that have influenced their engineering design process. It details how students might assimilate inspirations gleaned from the showcased designs. Students often mentioned incorporating specific features from a preferred existing design (Idtn-1, Idtn -2). They also drew lessons from less favoured designs they did not select. They commented on how they can enhance the design and implement it in their future backpack design project (Ideation-3, Ideation-4). For instance, as seen in Ideation-4, when a backpack was perceived to have limited space, a student suggested it needed more capacity, thus will secure enough space for their own backpack design.

Insights (Inst): Reasoning for selecting and not selecting

- **Inst-1:** "…backpacks are way more comfortable and better in ALL ways than a purse; A backpack will lay on your shoulders with straps, while a purse will irritate the skin on the inside of your elbow…"
- **Inst-2:** "A shows a backpack with only rolling capabilities while B shows multiple pockets which lets us know it can store multiple things inside of it."

As seen in the examples, students derived insights from both their chosen design and the one they did not select (Inst-1). They frequently contrasted features present in one design with its strengths, but absent or weakly displayed in the other. Such reasoning amplifies the advantages of comparing two competitive designs. Moreover, when both designs appear equally compelling, students sometimes employ user-centred perspectives as the decisive evaluation criterion (Inst-2).

Insights (Inst): Detailed comment with many criteria

- **Inst-3:** *"…looks like a good design, A backpack meant for travel, like a suitcase, but at the airport you can use it like a backpack. … It shows that you can use the strap and wheel to carry it like a travel trunk if it is heavy."*
- **Inst-4:** *"…this is more effective because you can carry more items and you have no weight on your back".*
- **Inst-5:** *"…you can adjust its length … it is more heavy duty and can also carry a lot of stuff…"*

Quality argumentation involving insights demonstrates a consideration of multiple, logically detailed criteria, indicating that students have thoughtfully evaluated various aspects of the design. For example, in learning the key features required for an effective backpack design, students show an understanding from Inst-3 to Inst-5. This approach, in contrast to brief and dismissive comments lacking substantial reasoning, highlights the students' comprehensive consideration of the multifaceted features of the better designs.

Insights (Inst): Explanation of aesthetic features

- **Inst-6:** *"…looks very impractical, but somewhat stylish."*
- **Inst-7:** *"…looks cool and that looks cool to the baby."*
- **Inst-8:** *"…like this one better because it has more colour and design."*

In earlier sub-codes, the term 'design' predominantly referred to the 'functionality' of the backpack, focusing on how it operates and serves practical purposes. However, the current sub-code 'Insights: Explanation of aesthetic features' shifts this focus, underscoring the aesthetic aspect of the backpack's design as a crucial characteristic, distinct from its functionality. Here, 'design' is interpreted in terms of the backpack's visual appeal and style ('how it will look'), rather than its functional efficiency or practical aspects ('how it will work'). Inst-6 shows how students prefer visual aspects over its practical aspects.

This emphasis on aesthetics involves considering elements that contribute to the backpack's visual attractiveness, artistic quality, and overall beauty (see Inst-7 and Inst-8). Such elements are integral to engineering design, going beyond mere functionality. By acknowledging the importance of aesthetics, we gain a deeper insight into the backpack design. This approach not only fulfils the basic functional requirements of the backpack but also delves into a more comprehensive understanding of user preferences and the product's appeal in the marketplace. Recognizing the aesthetic dimension is crucial as it can significantly influence user satisfaction and the product's success in a competitive market.

Empathy and Insights (Empt/ Inst): Explanation of cultural awareness

- **Empt/ Inst-1:** "Option B is widely used in many traditional cultures, but seems less efficient and looks unhealthy compared to modern backpacks."
- **Empt/ Inst-2:** "…lady carrying straw on her head and it would work if you don't have anything to carry an item…"

The backpack designs studied incorporate item-carrying methods from diverse cultural backgrounds. We value argumentation that demonstrates an understanding of these cultures and draws insightful interpretations from the designs. Specifically, such argumentations reflect an awareness and appreciation of various cultural practices and values (refer to Empt/ Inst-1 and Empt/ Inst-2). While students did not always prioritize traditional, ethnic carrying methods, they often displayed an understanding of the cultural context in which the backpack is used. Recognizing these cultural elements is considered valuable argumentations (Roth & Lee, 2004; Seah, 2005). It broadens students' cultural and user perspectives and promotes a more inclusive approach to product design, particularly for the global market.

Poor-quality Argumentation

Codes signifying poor-quality argumentation often exhibit deficiencies in empathy, ideation, and user insights, lacking essential argumentation components. Certain codes present little to no discernible argumentation due to their brief and dismissive nature, while others show biased or incorrect argumentation. Furthermore, as observed, brief and dismissive responses are prevalent across various codes. This means the same argumentation often simultaneously fall into three distinct categories: Vague, uncertain, and lack of knowledge.

Vague (Vg)

- **Vg-1**: "Cuz"
- **Vg-2**: "Mmm"
- **Vg-3**: "No no bag"

Argumentation deemed vague is considered as poor-quality argumentation because it leaves researchers uncertain about the comment's intent concerning the design. Argumentation seen in the examples (Vg-1 and Vg-2) are ambiguous and, therefore, categorized as vague. Additionally, ungrammatical remarks (refer to Vg-3) are also considered vague.

Uncertainty/ Lack of knowledge (Unct/ Lk)

- **Unct/Lk-1**: *"idk"*
- **Unct/Lk-2**: *"not sure"*

At times, students were uncertain or lacked the requisite experience or knowledge to express their argumentation. As a result, they offered argumentation like Unct/Lk-1 (an abbreviation for "I don't know") or Unct/Lk-2. These were labelled as 'uncertainty' or 'lack of knowledge.' Given that these argumentations do not provide even basic reasoning, they were classified as bad argumentation.

Too Short Response (TSR)

- **TSR-1:** "much space."
- **TSR-2:** "don't close."
- **TSR-3:** "better color."

While the argumentation was comprehensible, when the comment is too short in length, it lefts the argumentation underdeveloped (see TSR-1 to TSR-3). For example, the comment TSR-1 could imply that the backpack doesn't close properly, which might lead to issues such as items spilling out during movement. However, since these specifics weren't provided in the argumentation, they were categorized as insufficient argumentation.

Inappropriate Answer (IA)

- **IA-1:** "Looks just dumb."
- **IA-2:** "Bad design."
- **IA-3:** "Wrong design."

As seen in the examples above (IA-1 to IA-3), argumentation that are inappropriate, accusatory, or fault-finding were categorized as being of poor quality. It was coded in such a way because those argumentations do not offer valuable empathy, ideation, or insights in their argumentation.

Off-topic (Oft)

- **Oft-1:** "My mom says I am so special!"
- **Oft-2:** "Ride"

Sometimes, argumentation was completely irrelevant to the backpack design context, as demonstrated in the provided examples. These remarks were also classified as poor quality.

Gender Biased (GB)

- **GB-1:** "Who would use a girly laced bag as a way of carrying your stuff"
- **GB-2:** "Pink is just for girls"

A couple of designs featured lacy details and pink/red hues. As seen in the examples, some argumentation made gender-biased remarks, suggesting that such designs were solely for girls. Due to their inappropriate and gender-stereotyped nature, these argumentations were classified as poor-quality argumentation.

Cultural Stereotyping (CS)

- **CS-1:** "Products from [country A] are just bad."
- **CS-2:** "Designs from [country B] are cool."

Several designs displayed bags with culturally specific characteristics, illustrating diverse ways of transporting items. Certain explanations either continued or encouraged stereotypes linked to particular cultures or ethnicities. For instance, when a teacher presented a revised image of a backpack that evoked specific cultural associations (refer to Figure 6, which was substituted due to licensing constraints), it led students to associate the backpack with the negative or positive stereotypes they previously held. At times, these argumentations involved negative stereotypes targeting a particular culture (e.g., CS-1) or uncritically praising another (e.g., CS-2).

Figure 7. Images incorporated by teachers highlighting cultural stereotyping. OpenAI. (2023). Fictional backpack with fictional company logo [digital image]. DALL-E Image Generation.

Such argumentation was considered as poor-quality argumentation. It broadly categorizes designs based on limited examples, which can reinforce existing prejudices unless addressed and rectified.

Discussion

This study addresses the research question: "How is high-quality argumentation defined and operationalized within the context of engineering design thinking, and conversely, what characterizes poor-quality argumentation in this field?" The findings provide a detailed framework for understanding student argumentation, identifying 13 distinct codes classified into two overarching categories: 'Quality argumentation' and 'Poor-quality argumentation.' These classifications directly respond to the research question by operationalizing argumentation quality and providing nuanced characteristics of both high- and poor-quality argumentation.

The delineation of 'Quality argumentation', characterized by empathy, ideation, and insights, illustrates how high-quality argumentation manifests in user-centred design thinking. This aligns with the study's objective of defining exemplary argumentation within engineering education (Dym et al., 2005). Specifically, empathy emerged as a critical attribute, with student comments demonstrating user-centred perspectives that addressed the needs and contexts of end-users. Similarly, ideation subcodes revealed creative and analytical approaches to improving design, while insights highlighted the evaluative skills essential for engineering judgment. These findings define high-quality argumentation as a combination of critical thinking, creativity, and detailed reasoning, addressing the first part of the research question.

Conversely, 'Poor-quality argumentation' was marked by vagueness, brevity, biases, and a lack of understanding of the design context. These attributes respond to the second part of the research question by characterizing poor-quality argumentation as the absence of depth, clarity, and inclusivity. The overlap of codes, such as vagueness and brevity, suggests a correlation between unclear understanding and weak articulation, emphasizing the need for explicit teaching practices and exposure to models of high-quality argumentation. Additionally, the presence of gender biases and cultural stereotyping underscores the necessity of fostering culturally sensitive and inclusive argumentation practices within curricula.

The implications of these findings are significant for educators and curriculum designers. To address the research question's focus on operationalizing high-quality argumentation, the study highlights the need to provide students with clear benchmarks and structured opportunities to practice argumentation skills. For poor-quality argumentation, targeted interventions are required to reduce its occurrence by explicitly teaching critical reasoning and offering corrective feedback.

The study’s findings further advocate for integrating empathy, creativity, and cultural awareness into engineering education. These results suggest that instructional strategies should emphasize not only the technical aspects of engineering design but also the human and contextual elements that define quality argumentation. For curriculum designers, this means developing resources that clearly exemplify high-quality argumentation while addressing potential biases and stereotyping.

Finally, the limitations of this study, including the small sample size and the contextual specificity of the engineering design journals—shaped by the educational level, curriculum design, and instructional methods—must be acknowledged. While these constraints limit the generalizability of the findings, they also open pathways for future research. Studies that examine broader populations and varied design contexts could deepen our understanding of how argumentation quality evolves and is influenced by different educational settings.

In summary, this study directly addresses the research question by defining and operationalizing high-quality argumentation while identifying the key characteristics of poor-quality argumentation. Its contributions to engineering education highlight the pedagogical importance of argumentation and provide a foundation for future research aimed at enhancing critical thinking and reasoning in engineering design.

References

Andersson, E. K., Sjöström-Strand, A., Willman, A., & Borglin, G. (2015). Registered nurses views of caring in coronary care – a deductive and inductive content analysis. *Journal of Clinical Nursing*, *24*(23–24), 3481–3493. https://doi.org/10.1111/jocn.12975

Artiles, A. J., Kozleski, E., Harry, B., Zion, S., & Tate, W. (2005). Addressing the Disproportionate Representation of Culturally and Linguistically Diverse Students in Special Education through Culturally Responsive Educational Systems1 Janette K. Klingner, University of Colorado at Boulder. *Education Policy Analysis Archives*, *13*(38), 2.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang, Y., Li, L., & Wu, N. (2009). Learning and Scientific Reasoning. *Science*, *323*(5914), 586–587. https://doi.org/10.1126/science.1167740

Bartholomew, S. R. (2021). Investigating the impacts of differentiated stimulus materials in a learning by evaluating activity. *Mississippi Valley Technology Teacher Education Conference*.

Bartholomew, S. R., Ruesch, E. Y., Hartell, E., & Strimel, G. J. (2020). Identifying design values across countries through adaptive comparative judgment. *International Journal of*

Technology and Design Education, *30*(2), 321–347. https://doi.org/10.1007/s10798-019-09506-8

Bartholomew, S. R., & Strimel, G. J. (2018). Factors influencing student design decisions: Comparative judgement as a method for assessing open-ended design problems. *Design and Technology Education: An International Journal*, *23*(3), 10–24. https://ojs.lboro.ac.uk/DATE/article/view/2513

Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2019). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education*, *29*(2), 363–385. https://doi.org/10.1007/s10798-018-9442-7

Cain, C. C. (2012). Underrepresented groups in gender and STEM: The case of black males in CISE. *Proceedings of the 50th Annual Conference on Computers and People Research*, 97–102.

Cross, N. (1982). Designerly ways of knowing. *Design Studies*, *3*(4), 221–227. https://doi.org/10.1016/0142-694X(82)90040-0

Dam, R. F. (2023, October 16). *The 5 Stages in the Design Thinking Process*. The Interaction Design Foundation. https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process

DeKalb County School District. (2021). *DeKalb County School District Demographics (2021-2022 School Year)*. https://www.dekalbschoolsga.org/about/

Díaz, C., Dorner, B., Hussmann, H., & Strijbos, J.-W. (2023). Conceptual review on scientific reasoning and scientific thinking. *Current Psychology*, *42*(6), 4313–4325. https://doi.org/10.1007/s12144-021-01786-5

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., & Klemmer, S. R. (2009). The effect of parallel prototyping on design performance, learning, and self-efficacy. *Stanford Technical Report*, *10*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1926ad304265f410c1f3b7419546e3edf3fde347

Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*. McGraw-Hill Education (UK).

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Duschl, R. A., & Osborne, J. (2002). Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education*, *38*(1), 39–72. https://doi.org/10.1080/03057260208560187

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education*, *94*(1), 103–120.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Erduran, S., & Jiménez-Aleixandre, M. P. (2007). *Argumentation in Science Education: Perspectives from Classroom-Based Research*. Springer.

Han, J. (2013). *Scientific reasoning: Research, development, and assessment* [PhD Thesis, The Ohio State University]. https://rave.ohiolink.edu/etdc/view?acc_num=osu1366204433

Honey, M., Pearson, G., & Schweingruber, H. (2014). STEM integration in K-12 education: Status, prospects, and an agenda for research. In *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

IDEO. (2013). *Design thinking for educators*. IDEO. https://page.ideo.com/design-thinking-edu-toolkit

Jackson, A., Mentzer, N., Bartholomew, S. R., Lee, W., Yauney, J. M., Thorne, S., & Bayah, D. (2023, June 25). *Board 332: Learning by Evaluating (LbE): Engaging Students in Evaluation as a Pedagogical Strategy to Improve Design Thinking*. 2023 ASEE Annual Conference & Exposition. https://peer.asee.org/board-332-learning-by-evaluating-lbe-engaging-students-in-evaluation-as-a-pedagogical-strategy-to-improve-design-thinking

Jonassen, D. H. (2011). Supporting argumentation in problem solving. *Educational Technology Research and Development*, *59*(3), 439–457. https://doi.org/10.1007/s11423-011-9207-1

Karabiyik, T., Magana, A. J., Parsons, P., & Seah, Y. Y. (2023). Pedagogical approaches for eliciting students' design thinking strategies: Tell-and-practice vs. contrasting cases. *International Journal of Technology and Design Education*, *33*(3), 1087–1119. https://doi.org/10.1007/s10798-022-09757-y

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Keogh, B., & Naylor, S. (2000). Teaching and learning in science using concept cartoons: Why Dennis wants to stay in at playtime. *Investigating*, *16*(3), 10–14. https://search.informit.org/doi/abs/10.3316/aeipt.104804

Kimbell, R. (2022). Examining the reliability of Adaptive Comparative Judgement (ACJ) as an assessment tool in educational settings. *International Journal of Technology and Design Education*, *32*(3), 1515–1529. https://doi.org/10.1007/s10798-021-09654-w

Kimbell, R., & Stables, K. (2007). *Researching Design Learning: Issues and Findings from Two Decades of Research and Development*. Springer Science & Business Media.

Krell, M., Dawborn-Gundlach, M., & van Driel, J. (2020). Scientific reasoning competencies in science teaching. *Teaching Science*, *66*(2), 32–42.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

McNeill, K. L., & Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning. In J. Luft, R. Bell, & J. Gess-Newsome (Eds.), *Science as inquiry in the secondary setting* (pp. 121–134). National Science Teaching Association.

Mentzer, N., Lee, W., & Bartholomew, S. R. (2021). Examining the Validity of Adaptive Comparative Judgment for Peer Evaluation in a Design Thinking Course. *Frontiers in Education*, *6*, 1–15.

Mentzer, N., Lee, W., Jackson, A., & Bartholomew, S. (2023). Learning by Evaluating (LbE): Promoting meaningful reasoning in the context of engineering design thinking using Adaptive Comparative Judgment (ACJ). *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-023-09853-7

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020. https://doi.org/10.1002/tea.20035

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 281–300.

Pollitt, A. (2015). On 'Reliability' bias in ACJ. *Cambridge Exam Research*, *10*, 1–9.

Roth, W.-M., & Lee, S. (2004). Science education as/for participation in the community. *Science Education*, *88*(2), 263–291. https://doi.org/10.1002/sce.10113

Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition and Instruction*, *16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Seah, W. T. (2005). Negotiating about perceived value differences in mathematics teaching: The case of immigrant teachers in Australia. *Teaching and Teacher Education*, *21*(1), 63–80. https://doi.org/10.1016/j.tate.2004.11.001

Settlage, J., & Sabik, C. M. (1997). Harnessing the positive energy of conflict in science teaching. *Theory Into Practice*, *36*(1), 39–45. https://doi.org/10.1080/00405849709543743

Slavit, D., Grace, E., & Lesseig, K. (2021). Student ways of thinking in STEM contexts: A focus on claim making and reasoning. *School Science and Mathematics*, *121*(8), 466–480.

Solomon, J., Duveen, J., Scot, L., & McCarthy, S. (1992). Teaching about the nature of science through history: Action research in the classroom. *Journal of Research in Science Teaching*, *29*(4), 409–421.

Strimel, G. J., Bartholomew, S. R., Purzer, S., Zhang, L., & Ruesch, E. Y. (2021). Informing engineering design through adaptive comparative judgment. *European Journal of Engineering Education*, *46*(2), 227–246.

Thorne, S., Mentzer, N., Bartholomew, S., Strimel, G. J., & Ware, J. (2024). Learning by evaluating as an interview primer to inform design thinking. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-024-09933-2

Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.

Wallon, R. C., Jasti, C., Lauren, H. Z. G., & Hug, B. (2018). Implementation of a Curriculum-Integrated Computer Game for Introducing Scientific Argumentation. *Journal of Science Education and Technology*, *27*(3), 236–247. https://doi.org/10.1007/s10956-017-9720-2