ASME Journal of Mechanical Design
Online journal at:
https://asmedigitalcollection.asme.org/mechanicaldesign

Paper or Tablet? The Impact of Digital Tools on Sketching During Engineering Design Concept Generation

Madhurima Das[1]
Mem. ASME
Department of Mechanical Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139
e-mails: rimadas@mit.edu;
rimadas@alum.mit.edu

May Huang
Department of Mechanical Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139
e-mail: mayhuang@mit.edu

Maria C. Yang
Mem. ASME
Department of Mechanical Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139
e-mail: mcyang@mit.edu

Sketching is an important tool for engineers during concept generation. Sketch quantity during this early stage of design has been linked with eventual design outcomes and sketch quality has been linked with design perceptions. As such, both are important metrics to track. Prior work has also found gender differences in some of these sketch attributes, and considering gender imbalances in the field, it is important to assess whether there are differences in performance by gender during concept generation and if the use of digital tools exacerbate or mitigate any of these potential differences in performance by gender. Given the increase in use of digital tablets for sketching, it is important to assess how tablet use affects early stage engineering design sketches. This is especially important as tablets can automatically smooth lines and help perfect sketch features, but these features may also take longer to use and may result in fewer sketches produced. This study investigates differences in sketch quality, quantity, and understandability (the effectiveness of the sketch as a communication tool) between sketching on a tablet and sketching with pen on paper during an engineering design concept generation exercise. Results indicate that there is no difference in sketch quantity or understandability between the two tools. However, sketch quality, smoothness, and proportion/accuracy are all higher for the pen and paper condition than for the tablet condition. Finally, no gender differences in performance for either sketch quantity or quality were found. [DOI: 10.1115/1.4065458]

Keywords: creativity and concept generation, design methodology, design representation, design theory, design theory and methodology, design visualization, ethics and design, product design

1 Introduction

Sketching is a crucial part of early stage engineering design as a way to visually represent concepts and help designers think about their designs [1,2]. The majority (70–80%) of design life cycle costs have been estimated to be determined by decisions made in preliminary stages of design [3]. As such, changes to the early stages of the design process can impact the overall design performance and cost.

Another important aspect of sketching to investigate includes potential differences in perceived sketch performance between different groups of designers. Prior works have found that there may be gender differences in sketching performance in terms of sketch quantity and quality [4,5], which also has important implications for gender equity in design. Similarly, prior work has found that certain interventions during the design process can trigger "gender faultlines" and result in a negative impact on the number and creativity of ideas generated during concept generation [6]. It is imperative to understand whether the use of digital tools could potentially serve as a trigger for these faultlines or otherwise influence the concept generation process in a gendered fashion.

The focus of this study is concept generation sketching, specifically for engineering design problems. At this stage of the process, technical feasibility is not the focus—rather, it is important for engineers and designers to focus on creativity and having a large quantity of divergent ideas [7] and use sketches to communicate details of their designs [8]. Prior works have shown that sketch quantity (rather than quality) during concept generation is linked with better design outcomes by the end of the process [9–13], potentially due to having a wider variety of initial concepts to choose from when selecting and combining ideas throughout the design process. On the other hand, sketch quality is also important to measure as higher quality sketches are often perceived as being more creative even when representing the same design concept [14,15], so sketch quality is especially important in team design settings where others are making judgments about the creativity of a design concept. Two key metrics, line smoothness and proportion/accuracy, have consistently been identified as being part of sketch quality [5,9,16]. As such, these are the aspects of sketch quality studied in this context. Additional research has shown that

[1]Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received September 2, 2023; final manuscript received April 18, 2024; published online May 21, 2024. Assoc. Editor: Jessica Menold.

sketch quality and idea quantity are independent measures [17]. As a result, this study focuses on both the quality and quantity of sketches produced.

1.1 Motivation. Digital design tools have become increasingly common, especially as remote work has become more of the norm in both education and industry. In particular, tablets have become popular for notetaking and sketching in a streamlined fashion without having to worry about carrying the right notebook, running out of ink, or running out of pages to write on. These tools are new enough to the design process that they have not yet been studied in depth. We seek to build on our prior work investigating sketch quantity and quality between different tools for sketching [5]. The previous study was conducted where subjects were asked to sketch with pen, pencil, and tablet for 5-min high-level brainstorming sessions [5]. In this study, a new experiment was designed and conducted with the focus instead on sketching for more detailed concept generation during a longer time period specifically in the context of engineering design problems.

The motivations for conducting this research span from *financial motivations* of understanding the impacts of these new technologies since 70–80% of design life cycle costs come from decisions made in these early design stages [3], *social motivations* of determining if there are gender differences in performance between different tools since some changes to the design process can trigger "gender faultlines" and negatively influence results of concept generation [6], and *ethical motivations* of assessing how differences in access to digital drawing technologies may impact design performance and perception due to the increased cost and barrier to access of digital design technologies [18,19].

1.2 Summary. An experiment was conducted involving 40 participants who generated concepts by sketching ideas for two engineering design prompts, one on paper with a pen and one on a tablet (iPad) using a stylus (Apple Pencil). All participants completed both prompts in order to ensure that they did both tablet and paper sketching. Subsequently, the sketches were analyzed for quantity, quality, and understandability (effectiveness of communication) [9,16]. This work investigates the influence of the sketching tool used on sketching outcomes during engineering concept generation such as sketch quantity and quality. Additionally, it explores whether or not there are any gender differences in sketch quantity or quality during brainstorming for engineering problems.

The aim of this work is threefold: first, to explore whether or not the use of tablets influences the quality and quantity of sketches made during brainstorming, since this may influence eventual design outcomes. Second, to understand whether tablet use provides designers a "boost" in perceived sketch quality, which may have implications for equity depending on who is able to access the technology. Third, to investigate whether or not there are gender differences in any sketch characteristics for engineering design concept generation, with particular attention to whether or not any gender differences are exacerbated or minimized through use of tablets for sketching.

2 Related Work

2.1 Early Stage Concept Sketching. Sketching is a key tool during engineering design especially during the early stages of the design process [20–24]. Sketches are used in several ways during this stage and can be primarily categorized as "thinking sketches" or "talking sketches" [25]. Thinking sketches help a designer clarify their design intent and explore the design space while talking sketches are used to communicate their concepts to others [25]. Later stages of design can also use "prescriptive sketches" which are used for designers to communicate designs to others who are not involved in the design process [25]. Radcliffe and Lee have also identified three types of sketches for the concept development stage in engineering design and industrial design: functional, geometric, and pictorial [26]. Functional sketches express the functional relationship between the elements in a concept and are often symbolic rather than to scale; these can facilitate thinking and may be most in line with Goel's thinking sketches [2,26]. Geometric sketches are more focused on showing dimensions and accurate mechanisms; these are more representative of sketches used in later stages of design after ideation [26]. Pictorial sketches are drawn in perspective to demonstrate an overview of all functional and geometric information and are well suited for communicating designs to others, which may be more in line with the ethos of talking sketches, though they are also more suited for later stages of design after a concept is refined to the level of understanding what a full assembly would look like [26]. This work is focused on the earliest stages of design when designers are exploring a broad design space in an expansive way. As such, the sketches explored in this work are primarily thinking sketches, talking sketches, and pictorial sketches [25,26].

Prior work in interior design has identified three main sketching "profiles" from this stage as well: the realization sketching profile which places emphasis on an applicable solution, the learning sketching profile which emphasizes the given problem, and the designer/reflective sketching profile which is focused on the personal design process [27]. This also brings "written sketching" into the conversation around sketching during early stages, which can include word maps and annotations [27]. The realization sketching profile and the learning sketching profile are most aligned with the thinking and talking sketches most representative of this stage of the engineering design process and are in line with the types of sketches explored in this study.

There have been several efforts made to assess sketch quality during this stage using metrics such as line smoothness, perspective, proportion, accuracy, speed of strokes, and line start/stop locations [14,28,29]. Two sketch quality metrics, line smoothness and proportion/accuracy, have consistently appeared across many of these works and as a result have been used in prior works to create a measure of overall sketch quality [5,9,16]. An additional metric of sketch assessment called "understandability" has been developed in recent works to assess the effectiveness of sketches as a communication tool, which is a common use of sketches in engineering design [5,9,16,30]. As such, these are the three qualitative characteristics of sketches that are assessed and quantified for analysis in this study.

Additionally, producing a higher quantity of sketches during these stages has been linked with having better design outcomes in many prior works, though higher sketch quality does not appear to be linked with better design outcomes [9–13,31]. However, higher quality sketches may be perceived to be more creative than lower quality sketches of the same concept [14,15]. Since quality and quantity of sketches have both been demonstrated to have an impact on design perception and outcome and understandability of sketches is linked with a primary use case for sketching in engineering design, it is important to assess all three elements of sketches when evaluating new sketching tools.

2.2 Gender Differences in Sketching Performance During Ideation. Previous studies have found some gender differences in sketching performance in both sketch quantity [5] and sketch quality [4]. A study in industrial design found that sketches made by women were rated as lower quality than those made by men [4]. This is important as sketches perceived to be as lower quality may also be perceived to be less creative [14,15]. It has also been shown that designers' prior backgrounds in mechanical engineering or industrial design manifest in different types of sketches produced, so it is important to investigate this gender difference in sketching from an engineering design standpoint [32]. Another prior study found that art-based sketching interventions can help increase sketching self-efficacy for women in engineering [33].

Prior work has also found that gender and gender diversity of teams do not necessarily correlate with creativity of sketches produced [34], though activating "gender faultlines" during the ideation process by asking participants to perform a gender biased design task (such as designing a razor for men) can negatively impact the number of ideas produced and the creativity of ideas [6].

Prior work has also linked some sketch attributes such as line smoothness with confidence, so gender-related effects may occur due to differences in confidence in environments such as engineering and industrial design where women are a minority [4,35,36] as women in some Science, Technology, Engineering, and Math (STEM) programs have been found to have lower confidence even when they have the same mastery of skills [37] and are often perceived by experts as having lower confidence than men [38]. Another study found a gender difference in sketch quantity with women producing more sketches than men [5]. Though underlying causes for this difference were not identified, it points to an important area of further investigation, which is explored in this study.

2.3 Use of Digital Tools in Design. As the world transitions to becoming more digital, it is natural that digital tools are developed for aiding designers. Existing products include immersive design tools that may help with visual thinking, but many of these come with a high cost and impractical constraints such as requiring a large open space that designers can walk around safely in while tethered to a virtual reality headset [39]. Other prior work has found that despite the availability of digital design tools, many designers prefer to ideate with pen and paper [40]. More recently, tablets have become a popular tool for sketching digitally, especially as they allow users to carry a single product that can have multiple colors, allow for erasing, store files easily, and reduce paper use [41]. Some work has explored interfaces between tablet sketching and digital tools that can help synthesize ideas and generate novel concepts related to the ones drawn by the designer (such as a new mug design given several mug designs), which may be useful for design refinement stages once an initial solution space has been identified [42]. However, studies show that some digital design tools may inhibit designers if introduced too early. For instance, using computer aided design (CAD) early in a design process can result in design fixation and reduced creativity [43,44]. Sketching on a tablet does not require as much of a learning curve as CAD, but it can still modify elements of a sketch such as smoothing lines and perfecting shapes, which may make designs look more refined than they are in concept [45,46]. Since line smoothness is an important attribute of sketch quality metrics [9,14,28,29,35], it is possible that tablet sketches will be perceived as being higher quality (and thus more creative [14]) than those drawn on paper with pen. Similarly, tablets make it easy to erase and focus on perfecting drawings and adding detail, which may cause designers to focus on refining sketches rather than creating a large quantity of rough sketches. As a result, designers using tablets may produce fewer, higher quality sketches which may be perceived as more creative ideas [14,15]. This may have implications for engineering instruction and practice, especially in terms of assessing the equity of who has access to tablets. As such, this study focuses on understanding these potential differences in sketching performance between digital tablets and paper and pen.

2.4 Study Contribution. There is minimal prior work studying the impact of tablet use during early stage engineering design concept generation. Thus far, studies have explored how tablets used with a stylus can be used to provide real-time sketching feedback to students [47,48]. These studies have found that digital tools can be very useful in training novices on best practices for sketching and can help students build skills in perspective drawing. Additionally, they have found that completing this training on a tablet does not hinder students' sketching skill development even when they are tested on paper drawing skills [49]. Other studies have shown that designers using physical journals may make a different number of sketches and annotations compared to designers using "hybrid" journals with digital aspects, though the trend is not necessarily consistent over time as physical journals had more conceptual design stage sketches than hybrid journals in one year but hybrid journal had more sketches at all stages of design in another year [50]. Other studies have used tablets for tracing and practicing sketching of specific objects, though this is a very different use case from the type of sketching that typically occurs during concept generation [51,52].

Additionally, there have been few studies assessing gender differences in sketch quantity and quality for engineering design concept generation. One prior study has found that women had higher idea fluency (sketch quantity) than men for an engineering concept generation exercise [53]. Similarly, prior work in related fields suggests that there may be gender differences in sketching performance for both quantity and quality that should be investigated further [4,5].

The focus of this study is on quick "thinking" sketches [2,25] and not refined sketches from later stages of the design process. The focus is also on sketches made specifically in response to an engineering design problem. Overall, the goal of this study is to understand how tablet use and gender may influence the quantity, quality, and understandability (the effectiveness of the sketch as a communication tool) of sketches made during engineering design concept generation.

3 Research Questions

RQ1: What impact, if any, do digital sketching tools have on the quantity of ideas sketched during brainstorming for engineering design problems?

We hypothesize that participants using tablets will create fewer sketches than those using pens as they may spend more time refining their sketches using the tablet features, which may lead to an overall focus on quality over quantity of sketches when compared to those using paper and pen.

RQ2: What differences, if any, exist in sketch quality or understandability between digital and paper sketches for engineering design concept generation?

This question explores how the use of tablets influences sketch quality and the sketch's effectiveness as a communication tool (understandability) [54]. We hypothesize that sketches drawn on tablets will have higher sketch quality as a result of the tablet's built-in tools and features for line smoothing and sketch refinement.

RQ3: What are the gender differences, if any, in sketch quantity, quality, and understandability in engineering design concept generation?

This question aims to determine whether gender differences identified in prior literature [4,5] persist in engineering design concept generation through the metrics of sketch quantity and quality. This includes assessing whether there are differences in sketch quantity, quality, and understandability between genders in this context. It also investigates whether there are any gender differences by tool in order to determine whether tablets might mitigate or exacerbate the gender differences found in previous works.

To investigate these three research questions, we conducted a study of four mixed gender treatment groups who used each set of tools (pen and paper or tablet) to sketch concepts in response to two different design prompts. The order in which they used the tools and the order of the prompts was switched to have an equal number of people in each of the four treatment categories. We then analyzed differences in sketch quantity and quality across the different tools and gender groups.

4 Methods

4.1 Participant Recruitment. A total of 40 students from all levels (13 undergraduate and 27 graduate) were recruited for the study via emails sent to the Mechanical Engineering Department at a New England University. As a result, all participants had an affiliation with the Mechanical Engineering Department (38 Mechanical Engineering majors/1 Mechanical Engineering minors/1 in a different department but with a primary graduate research advisor in Mechanical Engineering) and the vast majority of participants had prior experience taking design courses. Many also had work experience in design related fields or had done personal projects related to design. Students were not recruited from a class as the study was not affiliated with a specific course in the department. All participants received a small "sketching kit" of four different pens in a canvas pouch as an incentive to participate in the study. The overall gender breakdown of the participants was 18 men and 22 women. One participant used both she/her pronouns and they/them pronouns. Their results are grouped with the she/her group due to the sample size.

4.2 Design Prompt Selection and Development. All participants were asked to perform an activity of sketching ideas for two different prompts, both of which have been used previously in design theory literature. The two prompts, the peanut sheller [55] and milk frother [56,57] design challenges, were selected for their relative understandability and established usage within previous design literature. The peanut sheller prompt, originating on ThinkCycle [58], asks participants to design a way to shell peanuts for low-resource settings. Customer needs include preventing damage to the peanuts and unavailability of electric outlets as a power source. This design problem has been used in the previous literature on design fixation and idea generation [55]. The milk frother prompt was developed at Pennsylvania State University and instructs participants to design a device to froth milk in a short amount of time and with minimal instruction. This prompt has been featured in studies on creativity and design novelty [56,57]. The full final prompts for both exercises are included in the Appendices.

The intention was to offer participants a prompt that would inspire relatively detailed mechanisms, especially in comparison to a preceding study which asked participants to brainstorm more theoretical ideas for products given only 5 min of time [5]. The peanut sheller prompt was piloted in 1–3 person group settings to find an ideal length of time per prompt and to garner feedback on the clarity of the prompt. It was found that 45 min was considered too much time, with participants running out of ideas within the first 15. Prior literature supports the idea that having a long time to ideate can be helpful for generating a larger quantity of ideas and more novel ideas [59] and that later ideas are often better than early ones in the ideation process [60]. As such, the overall time for the prompt was reduced from 45 min to 20 min so that it was still higher than the 15 min period after which people were running out of ideas.

Additionally, it was found that some participants, especially those who are allergic to peanuts, are not aware of the experience of manually shelling a peanut, so a visual aid of a peanut was included in the prompt to help with clarity of what a peanut shelled and unshelled looks like. Finally, two requirements were specified: (1) avoid damaging the peanut and (2) indicate a power source.

Next, the milk frother prompt was added to allow for observing how each individual's sketching performance changed between tools and prompts. For this prompt, two requirements were added: (1) the frother must froth milk in a short amount of time and (2) it must require minimal instruction to use. These two requirements were added in order to mirror the two customer needs provided in the peanut sheller prompt.

Participants were told explicitly to generate as many ideas as possible and given 20 min per prompt to encourage them to continue sketching ideas even after initial ideas and brainstorming had been exhausted. The length of each session was typically around 1 hour total (including 20 min per prompt for a total of 40 min of ideation) due to buffer time at the beginning and end of the session for debriefing participants on the tasks and asking them to fill out post-task surveys.

4.3 Treatment Types. All participants filled out an online pre-survey prior to starting the experiment and a post-task survey after completing the design exercises. This pre-survey included demographic questions along with questions about their prior expertise (on a scale of 0–10) with sketching using pen, pencil, and tablets. The post-task surveys included questions about the study experience such as which tool they preferred using, which prompt they preferred, and how good they thought their drawings were for each category.

Participants were divided randomly into four treatment groups (10 people per group). The order in which the participants sketched the prompts, as well as which tool they sketched with first, was varied based on their treatment group. After sketching for one prompt with one tool, the participant would then perform the other prompt with the other tool. The full experimental procedure is shown in Fig. 1. All participants in each session (doing the experiment at the same time in the same room) were in the same treatment group. The four treatment groups are shown in Table 1. The four groups were relatively evenly split by gender. Groups A, B, and C all had six women and four men. Group D had four women and six men. This gender split across groups was a result of the random distribution.

The four combinations were designated to ensure each participant worked with both tools and both prompts, as well as to see if there were any noticeable learning effects due to the order in which someone used the tools or completed the prompts.

Participants were asked in the pre-survey to share their prior experience with pen and paper sketching, tablet sketching, and tablet notetaking by rating their level of experience so that the researchers could ensure a relatively even distribution of experience

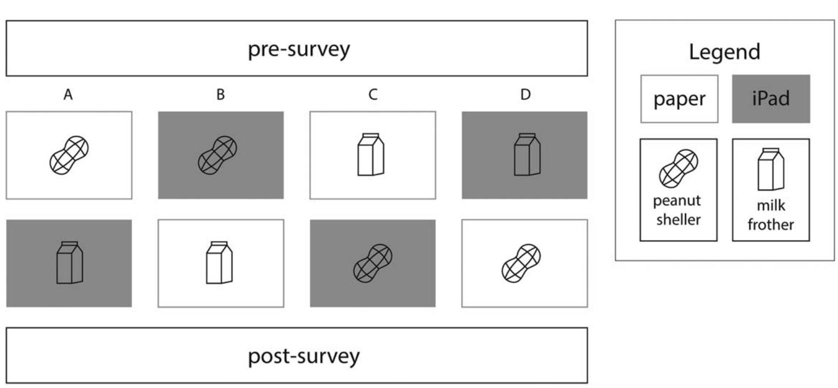

Fig. 1 Diagram of study methodology showing the order of surveys and prompts for each treatment group

Table 1 The four treatment groups used for the study

Group label	First prompt	Second prompt
A ($n = 10$, 6 women and 4 men)	Peanut sheller on paper	Milk frother on tablet
B ($n = 10$, 6 women and 4 men)	Peanut sheller on tablet	Milk frother on paper
C ($n = 10$, 6 women and 4 men)	Milk frother on paper	Peanut sheller on tablet
D ($n = 10$, 4 women and 6 men)	Milk frother on tablet	Peanut sheller on paper

levels across the four treatment groups. Ratings were done on a scale from 0 to 10 with 0 corresponding to having no prior experience with the medium and 10 corresponding to being very familiar with the medium. Zero was included as an option instead of starting the scale at 1 as it was possible that someone had never used a particular medium before (such as having never used a tablet for sketching). All four groups had relatively similar distributions of expertise as shown in Table 2. For all groups, the average scores were similar for pen and paper sketching and tablet notetaking. Similarly, for all four groups, participants had less experience, on average, with tablet sketching than they did with tablet notetaking and pen and paper sketching.

4.4 Experimental Setup. The experiment was conducted by two researchers. They gave participants consent documentation and once the consent forms were signed, the researcher read all instructions for the experiment aloud to participants from a prewritten script.

Tablet condition: Participants used an iPad with an Apple Pencil and the widely used writing and drawing app "Notability." The majority of participants already had experience with iPads and Notability (3 of 40 participants reported no experience sketching with tablets and 2 of 40 participants reported no experience taking notes on tablets). Prior experience with Notability is in part due to the fact that loaner iPads provided from the university to undergraduates and graduate Teaching Assistants (TAs) include Notability as a pre-loaded application. Participants did not receive any training using iPads prior to the experiment.

The application had been pre-loaded with the sketching template document (as seen in Fig. 2) with over 100 pages to ensure that participants would not run out of pages in the template during the time allotted. The participant's sketching process was recorded using the device's built-in screen recording feature. Participants were briefly told about the features of the Notability application including changing pencil width, erasing, changing colors, lasso tool (for copy/paste/resizing), undo/redo, and perfecting lines.

Pen and paper condition: Participants in the pen condition used a black Sharpie pen. Note that a Sharpie pen is different from a standard Sharpie marker. A Sharpie pen is a pen with a fine felt tip point that results in clear dark lines that do not smudge. In piloting the experiment, we tested pencils and a variety of pens such as ballpoint pens and ran into issues with ink/graphite smudging or being less visible in scans due to variable darkness. This is a particularly critical issue due to the line smoothness assessment of line quality that is used as a component of sketch quality. The Sharpie pen avoids these issues while maintaining a form that is the same as a standard pen or pencil and is similar to the size and shape of the Apple pencil. The paper used was white 8.5 in. × 11 in. letter-sized copy paper with the sketch template printed out.

For each drawing, participants were asked to write a short caption or idea name of less than a sentence to describe the concept in the spot provided on the template. Participants were informed that they could annotate the sketch if desired, but that they should try to communicate as much as possible through the sketch itself. At this stage, the researcher began screen recording on the tablet during the tablet condition. Finally, they were given the relevant concept generation prompt.

When the second round of 20 min had passed, they were asked to finish the last few strokes and put down their writing tool. Then, the researcher gave them the post-task survey and asked them to fill it out. At the end, participants were again reminded not to share information regarding the exercise with others since this was an ongoing experiment that ran over multiple weeks.

4.5 Sketch Quality Ratings. Sketches were rated using an adapted rubric and methodology developed and used in prior work on assessing early stage design sketches [5,9,16]. This prior methodology established line smoothness and sketch proportion/accuracy as the two subcomponents of sketch quality and is based on the metrics consistently used in the sketching literature [5,9,14,16,28,29]. It also introduced sketch understandability as an important metric to assess how well the sketch serves its purpose as a communication tool. The full rubric development methodology and justification for the metrics chosen can be found in Das and Yang's paper on Assessing Early Stage Design Sketches [16].

Representative sketches were selected from the dataset to serve as a rubric for raters. Sketches were rated using three independent scales of 1–5 for line smoothness, sketch proportion/accuracy, and concept understandability. For all cases, a score of 1 was the lowest score and 5 was the highest score. The full rubric can be seen in Fig. 3.

Instead of disregarding all text as was done in previous studies [5,9,16], the methodology was adapted to best replicate assessment

Table 2 Average experience scores (scale of 0–10) for each treatment group

	Pen and paper sketching	Tablet sketching	Tablet notetaking
A	5.80	3.90	5.80
B	6.40	4.90	6.40
C	5.70	3.80	6.30
D	5.40	3.90	6.00
Overall average	5.83	4.13	6.13

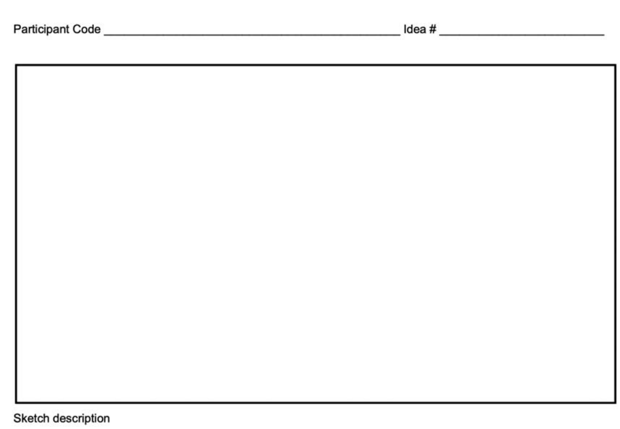

Fig. 2 Sketching template document

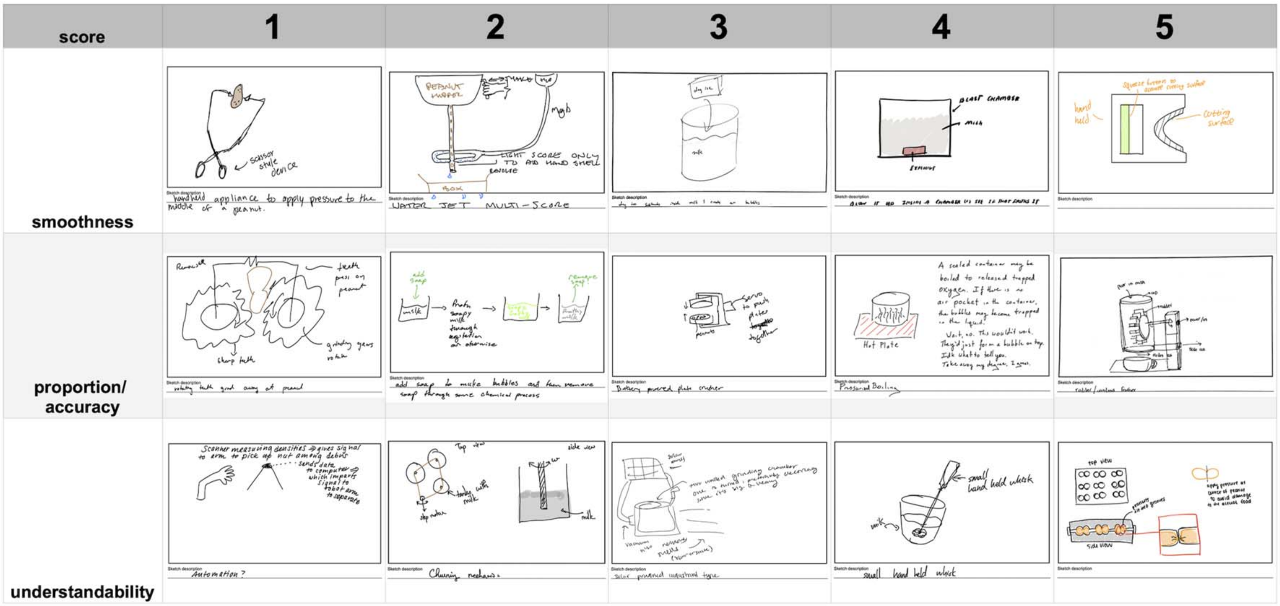

Fig. 3 Rubric with representative images from the dataset showing all score levels for each of the three scoring categories

techniques that would be used by practitioners in real settings. In engineering design settings, a primary goal of sketches is to communicate a concept, which is usually done using a combination of images and text. As such, the annotations (captions, labels, etc.) in images were not disregarded.

Two experts (a senior graduate student and a faculty member) worked together in order to establish "true" ratings by discussing each sketch as needed and completing the ratings collaboratively. A third rater then independently rated all the images in order to ensure that there was agreement between the two sets of ratings (set 1: raters 1 + 2 together, set 2: rater 3 independently).

All sketches were printed out individually on cardstock, with participant ID and sketch number removed. Raters then sorted sketches into five bins for each of the three sketch qualities, according to the rubric. All raters rated all sketches on smoothness first, then re-rated the randomized set of sketches on proportion/accuracy, and finally on understandability.

After an initial round of rating by both sets of raters, preliminary inter-rater reliability was calculated using Krippendorff's α (discussed in the results). If there was moderate or higher agreement within the category (based on guidelines from Ref. [61]), no further re-rating was done. If agreement was less than moderate, the absolute value of difference in scores between raters (from 0 to 4) was taken for each score. Sketches with a score difference of 2 or higher were rescored by the third rater (77 sketches in the smoothness category, 153 sketches in the proportion/accuracy category, and 86 sketches in the understandability category). Raters discussed these discrepancies to ensure that each rater held similar mental models of what the rubric meant and thus were assessing sketches consistently, though actual rating was conducted independently. Several areas were identified as areas of discrepancy in the method of rating. These are described here in order to assist with replicability of this rating method.

Overall, a score of 3 was considered to be the default or average. The final distribution of sketches was more similar to a normal distribution rather than being evenly distributed across the scale. An emphasis was made on using the full scale rather than a subset of the numbers to avoid a common pitfall of not using extreme ends of the rubric.

Smoothness focused on the overall smoothness of the majority of lines in the sketch rather than penalizing sketches if there was a minimal instance of feathering or an individual stroke that was less smooth. Smoothness of lines used in annotations (handwriting, arrows, etc.) was not considered. One important distinction made during this scoring process was in assessing pen strokes. Since the pen tip has a fixed width, occasionally when it was used for coloring or shading, the marks looked similar to feathering marks. Sketches that included this type of shading were not penalized.

For the understandability category, annotations within the sketch were considered when assessing understandability. The sketch description was not used as part of the rating, and instead was used to understand what the sketch intended to show and whether or not the sketch matched that intention. For this category, it was found that familiarity with objects presented seemed to influence these ratings. This is in line with previous work indicating that understandability is negatively correlated with concept novelty [5]. Importantly, technical feasibility of the concept was not assessed at this stage since these were initial concept generation sketches. In other words, "understandability" does not necessarily measure the ability to convey exactly how the concept will work. Rather, it assesses whether or not it is clear what the concept is.

Once all the ratings were completed with sufficient inter-rater reliability, the sketch ratings made by the two expert raters together were used for the sketch's final "true" score. Line smoothness and proportion/accuracy scores were added together for a total score out of 10 to represent the "overall sketch quality."

4.6 Sketch Quantity. The number of sketches made by each subject for each condition was counted. A small number of submissions contained only text, and thus were not counted as sketches. Additionally, pages used for diagrams or brainstorming documents such as mind maps were not counted.

5 Results

To help contextualize the results, four examples of sketches produced during the study are provided in Fig. 4 (one from each prompt on each tool). Each pair of images (C8a and C8b, D5a and D5b) was made by the same participant in order to show how the same person used each tool.

5.1 Statistical Analyses. The Wilcoxon rank sum test was used throughout to compare unpaired nonparametric scores such as the differences in sketching performance between genders.

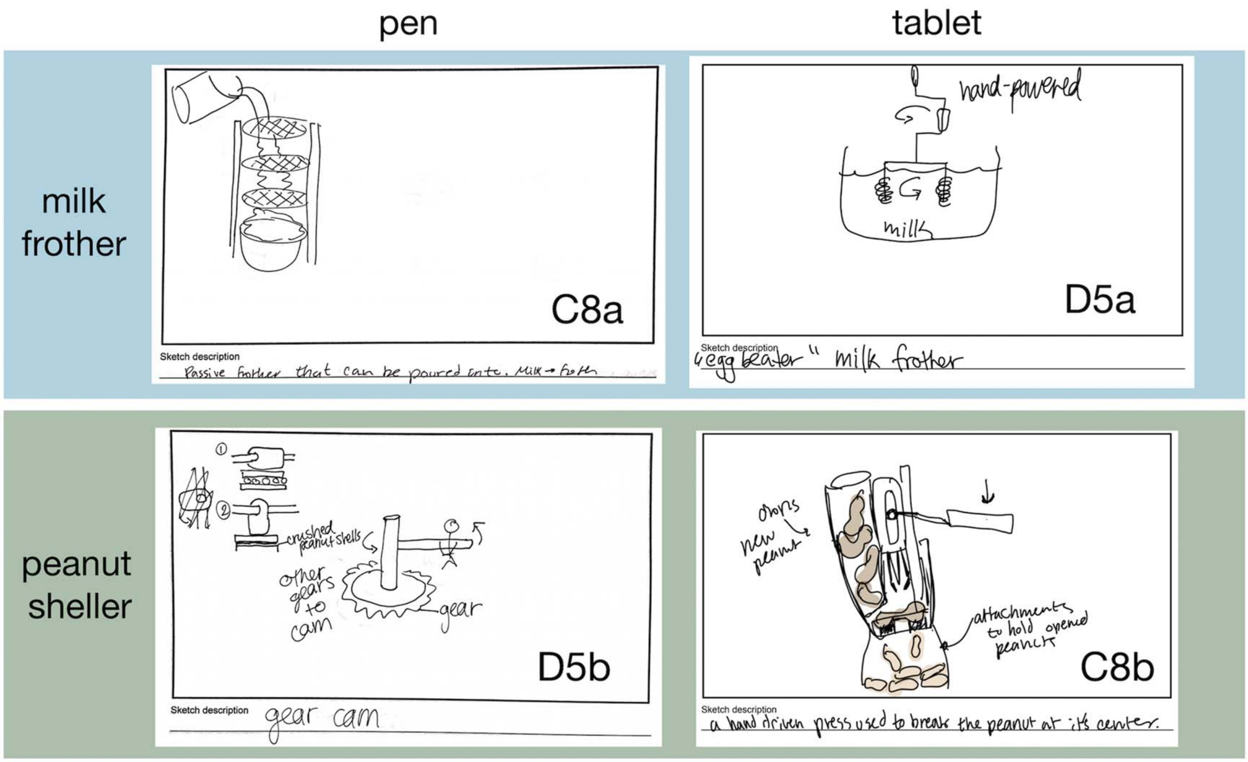

Fig. 4 Four sketches from the dataset. Sketches C8a and C8b are both made by participant C8 and the sketches D5a and D5b are both made by participant D5. Images C8a and D5a show milk frother sketches and D5b and C8b show peanut sheller sketches. Sketches C8a and D5b were made on paper and sketches D5a and C8b were made on a tablet. (C8a) Milk frother sketch with pen and paper, (D5a) milk frother sketch on tablet (D5b), peanut sheller sketch with pen and paper, and (C8b) peanut sheller sketch on tablet.

Similarly, the same test was used to assess the difference between sketch quality scores for each tool since the sketch quality scores are ordinal. The Wilcoxon signed rank test was used to compare paired nonparametric scores such as the differences between tablet and paper performance for each individual. Bartlett's test for equality of variances was used to check assumptions for both sketch quantity and sketch quality results and found that the variances were equal, so it was appropriate to proceed with the nonparametric tests used here.

Spearman's ρ was used to calculate correlations between nonparametric measures such as the correlation between sketch quality and prior experience using the relevant sketching tool for each individual. All statistics were computed at the 5% significance level ($p < 0.05$).

5.2 Potential Confounding Variables. In analyzing the sketch quantity results, the first step was to check for an ordering effect amongst the different conditions (more sketches in the first prompt and fewer sketches in the second prompt, or vice versa). This was checked using the Wilcoxon rank sum test for each type of drawing (peanut sheller on paper, peanut sheller on tablet, milk frother on tablet, and milk frother on paper) as well as overall using the paired Wilcoxon sign rank test to determine whether participants tended to create more sketches in their first prompt or second prompt. No statistically significant difference was found between the orders for each of these four types of drawings (peanut sheller on paper $Z = -1.7865$, $p = 0.0740$; peanut sheller on tablet $Z = 0.0597$, $p = 0.9524$; milk frother on tablet $Z = -1.4292$, $p = 0.1530$; milk frother on paper $Z = 0.4104$, $p =$ 0.0.6815). Similarly, no difference was found overall between first prompt and second prompt ($Z = -0.0760$, $p = 0.9395$). This is promising in terms of validating the methods of this study and allows us to compare the influence of the sketch tool overall without considering ordering as a confounding variable.

A statistically significant difference ($p < 0.01$) was found for sketch quantity between the two prompts, with the milk frother having one more sketch, on average, than the peanut sheller (mean of 7.525 versus mean of 6.400, respectively). The median and mean numbers of sketches produced for each prompt are shown in Table 3. This points to a need to further refine the prompts for any future studies that attempt to make comparisons between the two prompts in order to get similar results with both. However, within each prompt (peanut sheller or milk frother), there was no statistically significant difference between sketch quantity for the tablet and pen and paper conditions ($p = 0.7747$ for the peanut sheller prompt across tools and $p = 0.8169$ for the milk frother prompt across tools). Additionally, since an equal number of people did each prompt with each tool and all participants completed both prompts, we proceed with our overall analysis here.

5.3 Sketch Quantity. Overall, participants produced a total of 557 sketches with 287 created on the tablet and 270 on paper. The number of sketches produced by each participant ranged from 1 to 14 sketches on the tablet and 2 to 15 sketches on paper. The paired Wilcoxon sign rank test was used to determine whether participants

Table 3 Median and mean number of sketches produced for each ideation prompt

	Peanut sheller	Milk frother
Median	6	7
Mean	6.400	7.525

Table 4 Summary of mean scores and *p*-values for smoothness, proportion/accuracy, overall sketch quality, and understandability

	Pen and paper	Tablet	*p*-Value
Smoothness	3.12	2.82	$p<0.01$
Proportion/accuracy	3.55	3.21	$p<0.01$
Overall sketch quality	6.67	6.02	$p<0.01$
Understandability	3.54	3.51	0.87

tended to create more sketches on tablets than with pen and paper, or vice versa. No statistically significant difference ($Z=1.5570$, $p=0.1195$) was found in sketch quantity between the tablet (average of 7.175 sketches) and pen and paper (average of 6.75 sketches) conditions.

5.4 Sketch Quality. In order to assess inter-rater reliability between the raters' scores, Krippendorff's α was used since the data are ordinal (ordered but not necessarily equal intervals between categories) and was calculated using the open-source tool ReCal [62,63]. Krippendorff's α for agreement between raters was 0.618. This α value signals substantial agreement between the reviewers, as per Ref. [61].

Several differences in sketch quality were found between the pen and paper condition and the tablet condition and the results are summarized in Table 4. The Wilcoxon rank sum test was used to assess the difference between sketch quality scores for each tool since the sketch quality scores are ordinal. Sketch smoothness was statistically significantly different ($p<0.01$) between the two treatments. Pen and paper sketches had a mean smoothness score of 3.12 whereas tablet sketches had a mean smoothness score of 2.82.

The proportion/accuracy scores were also statistically significantly different ($p<0.01$) between the two treatments. Pen and paper sketches had a mean proportion/accuracy score of 3.55 whereas tablet sketches had a mean proportion/accuracy score of 3.21.

Overall sketch quality scores were also statistically significantly different ($p<0.01$) between the two treatments. Pen and paper sketches had a mean quality score of 6.67 whereas tablet sketches had a mean quality score of 6.02.

Sketch understandability scores were not statistically significantly different ($p=0.87$) between the two treatments (mean score of 3.54 for pen and paper and 3.51 for tablet). This may imply that both types of sketching mediums are effective communication tools.

5.5 Influence of Prior Experience on Sketch Quality and Quantity. Spearman's ρ was used to calculate correlations between participants' self-reported prior experience scores and their sketch quality scores for each sketching tool. This was explored in case it provided an explanation for the difference in sketch quality performance between tablet and pen sketches as outlined in Sec. 5.4.

Spearman's rank correlation was computed to assess the relationship between prior experience sketching with pen and paper and sketch quality. The relationship between prior experience sketching with pen and paper and sketch quality was not significant, $r(38)=-0.0060$, $p=0.9218$.

Spearman's rank correlation was computed to assess the relationship between prior experience sketching with tablets and sketch quality. The relationship between prior experience sketching with tablets and sketch quality was not significant, $r(38)=0.0360$, $p=0.5433$.

As such, prior experience with the tool does not explain the overall difference in sketch quality found between tablet and pen.

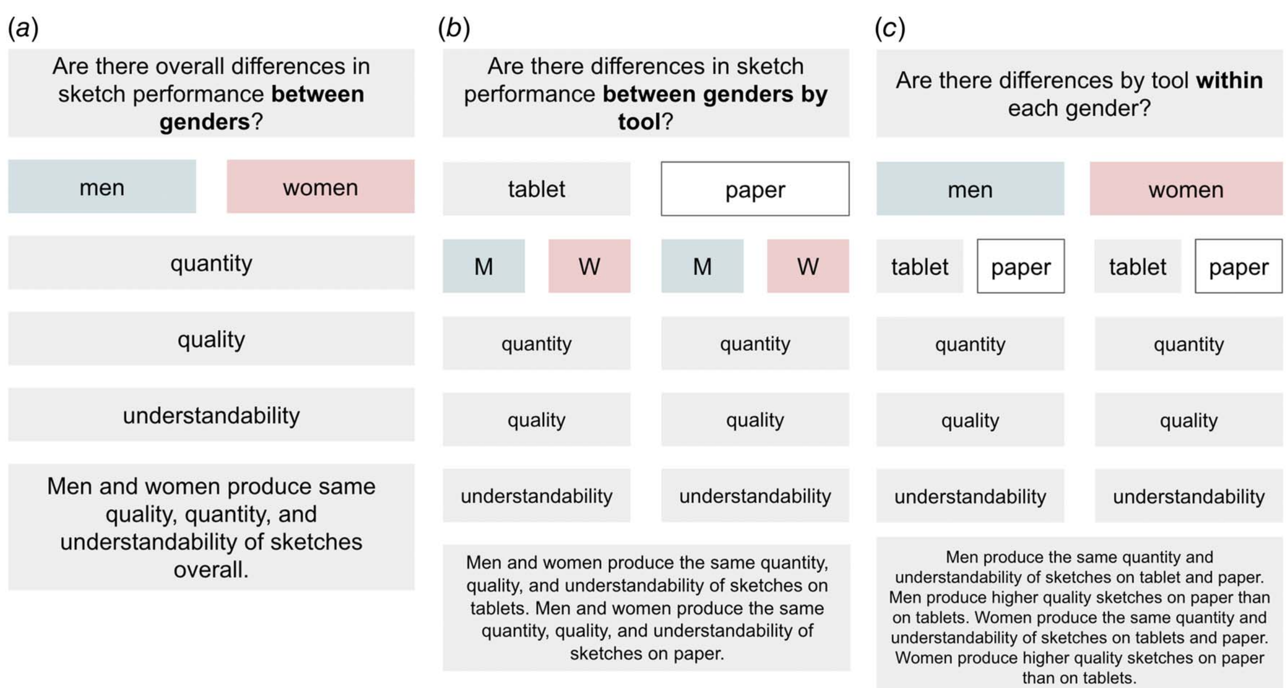

Fig. 5 Diagram representing the three analyses explored to determine whether there were any differences in sketch performance between genders, between genders by tool, or by tool within each gender. (*a*) Men and women produce the same quality, quantity, and understandability of sketches overall. (*b*) Men and women produce the same quantity, quality, and understandability of sketches on a tablet. Men and women produce the same quantity, quality, and understandability of sketches on paper. (c) Men produce the same quantity and understandability of sketches on tablet and paper. Men produce higher quality sketches on paper than on tablets. Women produce the same quantity and understandability of sketches on tablet and paper. Women produce higher quality sketches on paper than on tablets.

Table 5 Sketch metrics by gender (W is women, M is men)

Gender	Quantity		Smoothness		Proportion/accuracy		Understandability		Overall quality	
	W	M	W	M	W	M	W	M	W	M
Mean	14.90	12.70	2.94	3.00	3.37	3.38	3.54	3.52	6.31	6.38
Median	13	13	3	3	3	3	4	4	6	6
p-Value	0.4700		0.5277		0.8928		0.4553		0.8082	

We also investigated if prior experience with sketching tools had an impact on the quantity of sketches produced with each tool. This was of interest as there may have been a learning curve for people using a new tool such that they would spend more time adapting to the tool which may manifest in producing fewer sketches.

Spearman's rank correlation was computed to assess the relationship between prior experience sketching with pen and paper and sketch quantity. The relationship between prior experience sketching with pen and paper and sketch quantity was not significant, $r(38) = 0.3034$, $p = 0.0570$.

Spearman's rank correlation was computed to assess the relationship between prior experience sketching with tablets and sketch quantity. The relationship between prior experience sketching with tablets and sketch quantity was not significant, $r(38) = 0.0777$, $p = 0.6336$.

These results indicate that participants' prior experience with a sketching tool was not correlated with the quantity of sketches they produced.

5.6 Sketch Attributes by Gender. The Wilcoxon rank sum test was used to assess the difference between sketch attributes by gender of the overall population (not by treatment group).

This section examines three main questions regarding sketch performance by gender, each explored in a subsection below. Figure 5 shows a diagram representing the three sets of analyses conducted to explore the influence of gender on sketch performance by tool as well as the high-level overview of what effect was found.

5.6.1 Are There Overall Differences in Sketch Performance Between Genders?. There was no statistically significant difference in total sketch quantity between genders ($p = 0.47$). Men had an average of 12.7 sketches and women had an average of 14.9 sketches. Both had a median of 13 sketches. There was also no statistically significant difference in any sketch quality metric between genders. For line smoothness, women had an average score of 2.94 and men had an average of 3 ($p = 0.5277$). For proportion/accuracy, women had an average score of 3.37 and men had an average of 3.38 ($p = 0.8928$). For understandability, women had an average score of 3.54 and men had an average of 3.52 ($p = 0.4553$). For overall quality (smoothness + proportion/accuracy), women had an average score of 6.31 and men had an average of 6.38 ($p = 0.8082$). The full results are presented in Table 5 for clarity.

5.6.2 Are There Differences in Sketch Performance Between Genders by Tool?. There were no gender differences found in sketch performance between genders by tool, with results summarized in Table 6. There was no difference in sketch quantity between men (mean 6.06 sketches, median 6.5 sketches) and women (mean 7.32 sketches, median 7.5 sketches) using paper and pen ($p = 0.2613$) or men (mean 6.67 sketches, median 6.5 sketches) and women (mean 7.56 sketches, median 6.5 sketches) using a tablet ($p = 0.7015$). Similarly, there was no difference in sketch quality between men and women using paper and pen ($p = 0.3526$) or men and women using a tablet ($p = 0.2487$). Finally, there was no difference in understandability between men (average score 3.59) and women (average score 3.51) using paper and pen ($p = 0.7262$) or men (average score 3.45) and women (average score 3.57) using a tablet ($p = 0.1830$).

5.6.3 Are There Differences by Tool Within Each Gender?. The differences in sketch quality between tools observed at the population level persisted when the results were broken out by gender, as shown in Tables 7–9. There was no difference in sketch quantity scores for men using pen and paper versus men using tablets ($p = 0.5968$) or women using pen and paper versus women using tablets ($p = 0.8966$). This lack of difference in sketch quantity produced between tools is consistent with the result found at the overall population level. Overall sketch quality was higher for pen and paper than for tablets for both genders. For women, average sketch quality was 6.71 on paper and 5.92 on tablet with $p < 0.01$. For men, average sketch quality was 6.61 on paper and 6.16 on tablet with a $p < 0.01$. This difference in sketch quality between tools is consistent with the result found at the overall population level. There was also no difference in understandability scores for men using pen and paper versus men using tablets ($p = 0.2756$) or women using pen and paper versus women using tablets ($p = 0.5217$). This lack of difference in sketch understandability between tools is consistent with the result found at the overall population level.

In sum, there was no finding indicating a gender difference in performance overall or specific to a tool.

6 Discussion

Several of the results reported here were unexpected. Namely, the metrics that matter most for the design process and design outcomes, quantity of ideas and effectiveness of sketches for communication (understandability), are consistent between the two tools explored here. However, sketch quality overall is better with pen and paper than with the tablet. Further examining the subcomponents of the sketch quality score, both sketch smoothness and proportion/accuracy scores are higher on pen and paper than on the tablet. This is particularly interesting since the tablet software is designed to auto smooth lines and many participants additionally used the tablet's line straightening features. Similarly, it might be expected that proportion/accuracy scores would be higher on tablets since tablets can create "perfect" circles and rectangles. As such, it was expected that the tablet would afford users an advantage for both line smoothness and proportion/accuracy though this contradicts the result. Additionally, it is of note that these results are different from those in prior work that found no difference in sketch quality between pen and tablet for a shorter prompt that was less focused on engineering design [5]. Though we see this difference, it is unclear from this study why this difference exists. Turning to qualitative methods to better understand the experience of sketching with each tool could help elucidate the underlying reasons for these results. For instance, if the "feel" of the tablet is worse than that of paper, it may be more difficult to use effectively. Alternatively, it is possible that participants generally focused less on using some of these refining features during the study since they were completing a task within a specified time limit. A more open ended time frame may simulate the real-world ideation context better and show more nuanced results. The fact that pen and paper is producing the higher quality result is heartening as it does not point to digital tools artificially inflating sketching skill, which would pose concerns for equity and access to these more expensive tools. Instead, this points to the idea that the "default"

Table 6 Gender differences in sketch attributes by tool (W is women, M is men)

	Sketch quantity on paper		Sketch quantity on tablet		Sketch quality on paper		Sketch quality on tablet		Sketch understandability on paper		Sketch understandability on tablet	
	W	M	W	M	W	M	W	M	W	M	W	M
Mean	7.32	6.06	7.56	6.67	6.71	6.61	5.92	6.16	3.51	3.59	3.57	3.45
Median	7.5	6.5	6.5	6.5	7	6	6	6	4	4	4	4
p-Value	0.2613		0.7015		0.3526		0.2487		0.7262		0.1830	

Table 7 Sketch quantity differences split out by gender

	Pen and paper	Tablet	*p*-Value
Women	7.32	7.59	0.8966
Men	6.06	6.66	0.5968

Note: Sketch quantity was not different between the tools for either gender, mirroring the overall results.

Table 8 Sketch quality differences split out by gender

	Pen and paper	Tablet	*p*-Value
Women	6.71	5.82	$p<0.01$
Men	6.61	6.16	$p<0.01$

Note: Sketch quality was higher for pen and paper than for tablet for both genders, mirroring the overall results.

Table 9 Sketch understandability difference split out by gender

	Pen and paper	Tablet	*p*-Value
Women	3.51	3.57	$p=0.5217$
Men	3.59	3.45	$p=0.2756$

Note: Sketch understandability was not different between the tools for either gender, mirroring the overall results.

of pen and paper is not only acceptable but is resulting in higher quality output. As new digital tools are developed and marketed, there can be a desire to immediately incorporate them into the design flow, but these results indicate that the tools we are already using may actually be more effective, at least at this stage.

These findings are also exciting in terms of implications for tool accessibility for students from various socioeconomic statuses as they imply that students with access to digital tools are not necessarily producing better results in terms of sketch quantity, quality, and understandability. These findings may be of interest to design and engineering educators in terms of how they train novices to sketch for engineering design contexts. As some institutions move away from teaching skills in sketching and drafting on paper in favor of digital design skills, these results may suggest that sketching with pen and paper is still a better first step for skill building for novices. Due to the difference in sketch quality on paper and pen and tablet, educators working with team-based courses may want to ensure that students in teams are using the same medium for sketching so that sketches when compared to one another are not experiencing a difference in perceived creativity due to differences in sketch quality between the two tools [14,15].

In this work, no gender differences in sketch quantity or quality between different genders were found, which is promising for the field of engineering design. Related fields, such as industrial design, have found that women performed worse on sketching activities than men [4]. It is heartening to see that this earlier finding is not mirrored in this study in the engineering design space, despite similar discrepancies in representation in the field (19% of industrial designers are women [4] and 13% of mechanical engineers are women [64]). In particular, it is important to note that the findings indicate that the two genders studied here performed equally well when compared to one another. There was no finding indicating that one gender performed better on one tool than the other gender (e.g., women creating better tablet drawings and men creating better paper drawings). This is a very promising result as it indicates that there are no immediate serious concerns in terms of gender equity when using tablets for sketching if all else is equal in terms of access to the tools. However, this study only explores a few limited areas within the sketching experience, namely, quality, quantity, and understandability. It is possible that there are other elements of digital sketching, including qualitative experiences, that have more complex relationships with gender, and this should be explored further.

7 Limitations and Future Work

One main area of future work for this study is that there were some differences in sketch quantity scores between the two prompts. Further refinement of the prompts would help in ensuring that the two prompts are equivalent and could be used independently to produce similar results. Qualitative analysis of the post-task surveys may help clarify why this difference was present. Perhaps participants found the milk frother prompt to be easier to generate ideas for since they are more likely to have used an existing milk frother than an existing peanut sheller.

Additionally, no qualitative analysis of the sketching experience was done for this study. A common comment regarding tablet sketching is that the feel of the tablet and the type of tablet matters, especially if people have experience with a tablet other than the tablet. A qualitative analysis of the post-task surveys may help understand how the experience differed between the two sketching tools beyond the quantitative metrics investigated here and could help clarify the cause of the surprising quantitative differences in sketch quality found here. These preferences could vary between different disciplines as well, so future work could explore similar studies with other fields such as Architecture and Industrial Design. This was also a controlled study in a lab setting. It would be useful to study these tools in real contexts in order to determine if there are differences in how designers use each tool in practice.

It is important to acknowledge that these results may change as designers' levels of training or experience with each tool change, though level of prior experience was not correlated with sketch quality or quantity in this study. Perhaps initial tablet sketches are not as high in quality but with some training, designers can make better drawings on tablets than they can on paper. Similarly, these results must be contextualized within this specific phase of the design process. It is possible that differences between paper sketching and tablet sketching would be more pronounced at other stages of the process such as concept refinement. Additionally, there may be operational differences in terms of how both types of sketches can be used in team settings. The features of the tablet may provide benefits in terms of the types of drawings that can be produced but could result in more steps needed for use. For instance, tablet sketches may need to be printed in order to effectively show a variety of designs to a team at once and be able to compare them during a pinup session.

It is also possible that other stages of the design process would have different results. For instance, tablets may allow designers to quickly make more refined sketches during concept refinement phases. At this stage, it is possible that refined tablet sketches could be faster and more polished than sketches done by hand. The ability of tablets to make quick copies of and edits to sketches may also make it a more useful tool in team contexts where concepts are being iterated on rapidly by a group.

A potential confounding factor in this study with respect to gender is that the Mechanical Engineering Department at the institution studied here has a relatively higher proportion of women than the national average. At the undergraduate level, around 50% of students are women. At the graduate level, around 30% of the students are women. These are both much higher than the national average of 13% of mechanical engineers being women [64]. As a result, this university context may be an environment that is particularly welcoming to women and facilitates a space for women to thrive. This may contribute to the result of seeing no significant difference in sketch quantity and quality in engineering concept generation between students of different genders. Future work should explore similar studies at other institutions (both in academia and industry) that have different balances of genders present to see if the gender results found here change. If these results of no difference in performance by gender are truly based on the local context and gender ratio, this finding may be a data-driven way to demonstrate the value of having a more gender balanced department and warrants further study.

Future work regarding other aspects of this sketch study is in progress. These areas of future work include assessing qualitative differences in the sketching experience between the two tools based on responses to the post-task questions. Additionally, assessments of sketch novelty and concept evolution are being conducted to determine if there are differences between the two tools. Other work could include studying the role of digital and physical sketches in team settings in order to assess the sketches' effectiveness as communication tools (understandability) in their intended context.

8 Conclusions

Overall, this study finds that all else equal, tablets may be relatively acceptable alternatives to pen and paper sketching for early stage engineering design concept generation, but that there are some differences in sketch quality attributes between the two mediums that are important to keep in mind. Since there are no differences in the metrics that are most relevant for engineering design: sketch quantity (which correlates with long-term design outcomes) or the effectiveness of sketches as communication tools (understandability), which is often the main purpose of early stage engineering design sketches, tablet use may not pose a major risk in terms of influencing design outcome or team communication. However, the differences in sketch quality results between the two mediums with paper and pen sketches being higher quality than those made on the tablet should be kept in mind by designers considering using tablets for concept generation given the implications for perceived creativity of lower quality sketches. More detailed summaries of conclusions in response to each research question are presented below.

RQ1: What impact, if any, do digital sketching tools have on the quantity of ideas sketched during brainstorming for engineering design problems?

In this study, there was no statistically significant difference in sketch quantity between digital sketches and pen and paper sketches. This is different from expected results, as the expectation was that participants would spend more time refining sketches using the tablet's features, resulting in fewer overall sketches.

RQ2: What differences, if any, exist in sketch quality or understandability between digital and paper sketches for engineering design concept generation?

In this study, sketch quality was found to be higher on pen and paper than on tablets. Specifically, both smoothness and proportion/accuracy scores are higher for sketches made with pen than sketches made on tablets. However, there was no statistically significant difference in understandability scores between sketches made on tablets and sketches made with pen. This is also different from what may have been expected. Due to the tablet's built-in features for smoothing lines and refining sketches, it was expected that tablet sketches would be higher in quality than pen and paper sketches. The reason for this difference is not obvious and further qualitative analyses could help elucidate why this difference was observed.

RQ3: What are the gender differences, if any, in sketch quantity, quality, and understandability in engineering design concept generation?

In this study, there were no differences found in sketch quantity, quality, and understandability between students of different genders for engineering design concept generation. Furthermore, there were no gender differences by tool found. Additionally, the results within each gender matched the results from the overall population with both women and men having higher quality sketches on paper than on the tablet but no difference in sketch quantity or understandability between tools. These results are exciting because they point to differences in results for engineering design and industrial design. Prior work in industrial design found that women performed worse in sketching activities, but this is not the case in our engineering design context [4]. This is an especially notable finding because similar gender imbalances exist in engineering and industrial design [4,64] and because of the nuances of gender in STEM programs such as engineering where women have historically been outnumbered and had lower confidence even in situations where they have the same skill level [37]. The overall findings with respect to differences in the tools were mirrored when split by gender, which further reinforces the result that there were no differences in performance between the two genders investigated in this study. This is reassuring as it indicates that the introduction of tablets into design is not immediately posing an issue in terms of gender equity (as it would if one gender had superior performance on tablets and another performed better on paper).

Acknowledgment

We thank the participants in the study for their contribution to our work. A special thanks to Morgan Weaver for his advice on methods and best practices for sketch rating.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

Appendix A: Design Problem—Device to Froth Milk

Problem Description. Frothed milk is a pourable, virtually liquid foam that tastes rich and sweet. It is an ingredient in many coffee beverages, especially espresso-based coffee drinks (Lattes, Cappuccinos, and Mochas). Frothed milk is made by incorporating very small air bubbles throughout the entire body of the milk through some form of vigorous motion. As such, devices that froth milk can also be used in a number of other applications, such as for whipping cream, blending drinks, emulsifying salad dressing, and many others.

The goal of this project is to design and build *a new innovative product that froths milk in a short amount of time*.

Customer needs:

- Froths milk in a short amount of time
- Minimal instruction needed to use

Please see the reference image below for an example of frothed milk being poured into coffee.

You will have 20 min to brainstorm and sketch *as many ideas as possible*. The goal of this design task is not to produce a final solution to the design problem but to brainstorm ideas that could lead to a new solution. As such, focus on generating as many ideas as possible—*do not focus on the feasibility of your ideas at this stage*.

In your drawings, include enough detail such that someone could understand the main functions of your system. To clearly communicate your concepts, make your drawings large and easy to read with *one idea per sheet* using the template. Number your ideas in the designated location on the template. You can use as many sheets of paper as you like.

For each drawing, please *also write a short caption or idea name of less than a sentence to describe the concept*. You may annotate your sketches to help clarify the components of your designs.

Appendix B: Design Problem—Device to Shell Peanuts

Problem Description. In places like Haiti and certain West African countries, peanuts are a significant crop. Most peanut farmers shell their peanuts by hand, an inefficient and labor-intensive process. The goal of this project is to design and build a low-cost, easy to manufacture peanut shelling machine that will increase the productivity of the African peanut farmers.

Customer needs

- Remove the shell with minimal damage to the peanuts.
- Electrical outlets are not available as a power source.

Please see the reference image below for examples of a peanut in and out of its shell.

You will have 20 min to brainstorm and sketch *as many ideas as possible* for a peanut shelling machine. The goal of this design task is not to produce a final solution to the design problem but to brainstorm ideas that could lead to a new solution. As such, focus on generating as many ideas as possible—*do not focus on the feasibility of your ideas at this stage*.

In your drawings, include enough detail such that someone could understand the main functions of your system. To clearly communicate your concepts, make your drawings large and easy to read with *one idea per sheet* using the template. Number your ideas in the designated location on the template. You can use as many sheets of paper as you like.

For each drawing, please *also write a short caption or idea name of less than a sentence to describe the concept*. You may annotate your sketches to help clarify the components of your designs.

References

[1] Company, P., Contero, M., Varley, P., Aleixos, N., and Naya, F., 2009, "Computer-Aided Sketching as a Tool to Promote Innovation in the New Product Development Process," Comput. Ind., **60**(8), pp. 592–603.

[2] Goel, V., 1995, *Sketches of Thought*, MIT Press.

[3] Saravi, M., Newnes, L., Mileham, A. R., and Goh, Y. M., 2008, "Estimating Cost at the Conceptual Design Stage to Optimize Design in Terms of Performance and Cost," Collaborative Product and Service Life Cycle Management for a Sustainable World: Proceedings of the 15th ISPE International Conference on Concurrent Engineering (CE2008), London, UK, January, pp. 123–130.

[4] Barnhart, B., and Walters, K., 2018, "The Hot Industrial Design Sketch: Perpetuating the Dominance of the Male Industrial Designer," DS 93: Proceedings of the 20th International Conference on Engineering and Product Design Education (E&PDE 2018), London, UK, September.

[5] Das, M., Huang, M., and Yang, M. C., 2022, "Tablets, Pens, and Pencils: The Influence of Tools on Sketching in Early Stage Design," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, St. Louis, MO, Aug. 14–17.

[6] Pearsall, M. J., Ellis, A. P. J., and Evans, J. M., 2008, "Unlocking the Effects of Gender Faultlines on Team Creativity: Is Activation the Key?" J. Appl. Psychol., **93**(1), pp. 225–234.

[7] Toh, C. A., and Miller, S. R., 2015, "How Engineering Teams Select Design Concepts: A View Through the Lens of Creativity," Des. Stud., **38**, pp. 111–138.

[8] Westmoreland, S., Ruocco, A., and Schmidt, L., 2011, "Analysis of Capstone Design Reports: Visual Representations," ASME J. Mech. Des., **133**(5), p. 051010.

[9] Das, M., and Yang, M. C., 2021, "Assessing Early Stage Design Sketches and Reflections on Prototyping," International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Virtual, Online, August 2021.

[10] Yang, M. C., 2004, "An Examination of Prototyping and Design Outcome," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Salt Lake City, UT, Sept. 28–Oct. 2, Vol. 46962, pp. 497–502.

[11] Cham, J. G., and Yang, M. C., 2005, "Does Sketching Skill Relate to Good Design?" International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Long Beach, CA, September 2005, pp. 301–308..

[12] Song, S., and Agogino, A. M., 2004, "Insights on Designers' Sketching Activities in New Product Design Teams," Vol. 3a: 16th International Conference on Design Theory and Methodology, Salt Lake City, UT, Sept. 24–28, pp. 351–360.

[13] Takai, S., Esterman, M., and Midha, A., 2015, "An Approach to Study Associations Between Design Concepts and Design Outcomes," ASME J. Mech. Des., **137**(4), p. 041101.

[14] Kudrowitz, B., Te, P., and Wallace, D., 2012, "The Influence of Sketch Quality on Perception of Product-Idea Creativity," Artif. Intell. Eng. Des. Anal. Manuf., **26**(3), pp. 267–279.

[15] Kwon, J., and Kudrowitz, B., 2021, "The Sketch Quality Bias: Evaluating Descriptions of Product Ideas With and Without Visuals," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Anaheim, CA, Aug. 17–20.

[16] Das, M., and Yang, M., 2022, "Assessing Early Stage Design Sketches and Reflections on Prototyping," ASME J. Mech. Des., **144**(4), p. 041403.

[17] Weaver, M. B., Buck, J., Merzdorf, H., Dorozhkin, D., Douglas, K., and Linsey, J., 2022, "Investigating Priming Effects of Sketch Evaluation Instructions on Idea Generation Productivity," Volume 6: 34th International Conference on Design Theory and Methodology (DTM), St. Louis, MO, Aug. 14–17.

[18] McAlpine, H. C., Hicks, B., and Culley, S., 2009, "Improving Re-use of Informal Information Through the Creation of an Engineering Electronic Logbook (EEL): A Demonstrator," Vol. 2: 29th Computers and Information in Engineering Conference, Parts A B, San Diego, CA, Aug. 30–Sept. 2, pp. 843–850.

[19] Lohani, V., Castles, R., Lo, J., and Griffin, O., 2007, "Tablet Pc Applications in a Large Engineering Program," 2009 Annual Conference & Exposition, Honolulu, HI, June 14.

[20] Häggman, A., Tsai, G., Elsen, C., Honda, T., and Yang, M. C., 2015, "Connections Between the Design Tool, Design Attributes, and User Preferences in Early Stage Design," ASME J. Mech. Des., **137**(7), p. 071408.

[21] Kudrowitz, B. M., and Wallace, D., 2013, "Assessing the Quality of Ideas From Prolific, Early-Stage Product Ideation," J. Eng. Des., **24**(2), pp. 120–139.

[22] Yang, M. C., 2009, "Observations on Concept Generation and Sketching in Engineering Design," Res. Eng. Des., **20**(1), pp. 1–11.

[23] Yang, M. C., 2003, "Concept Generation and Sketching: Correlations With Design Outcome," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Chicago, IL, Sept. 13–16, pp. 829–834.

[24] Yang, M. C., and Cham, J. G., 2007, "An Analysis of Sketching Skill and Its Role in Early Stage Engineering Design," ASME J. Mech. Des., **129**(5), pp. 476–482.

[25] Ferguson, E. S., 1994, *Engineering and the Mind's Eye*, MIT Press.

[26] Radcliffe, D. F., and Lee, T. Y., 1990, "Models of Visual Thinking by Novice Designers," 2nd International Conference on Design Theory and Methodology, Chicago, IL, Sept. 16–19, pp. 145–152.

[27] Bar-Eli, S., 2013, "Sketching Profiles: Awareness to Individual Differences in Sketching as a Means of Enhancing Design Solution Development," Des. Stud., **34**(4), pp. 472–493.

[28] Hilton, E., Willifor, B., Li, W., McTigue, E., Hammond, T., and Linsey, J., 2016, "Consistently Evaluating Sketching Ability in Engineering Curriculum," Fourth International Conference on Design Creativity, Atlanta, GA, Nov. 2–4.

[29] Hammond, T., Kumar, S. P. A., Runyon, M., Cherian, J., Williford, B., Keshavabhotla, S., Valentine, S., Li, W., and Linsey, J., 2018, "It's Not Just About Accuracy: Metrics That Matter When Modeling Expert Sketching Ability," ACM Trans. Interact. Intell. Syst., **8**(3), pp. 1–47.

[30] Letting, C., Krishnakumar, S., Johnson, E., Zurita, N. S., and Menold, J., 2023, "Investigating the Effect of Sketch Quality on the Shared Understanding of Design Dyads," 24th International Conference on Engineering Design (ICED23), Bordeaux, France, July 24–28.

[31] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, A. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," ASME J. Mech. Des., **133**(3), p. 031008.

[32] Jia, M., Jiang, S., Hu, J., and Qi, J., 2023, "Toward Understanding Sources and Influences of Design Fixation: A Focus on Example Stimuli and Background of Novice Designers," ASME J. Mech. Des., **145**(5), p. 051402.

[33] Hu, W.-L., Booth, J., and Reid, T., 2015, "Reducing Sketch Inhibition During Concept Generation: Psychophysiological Evidence of the Effect of Interventions," Vol. 7: 27th International Conference on Design Theory and Methodology, Boston, MA, Aug. 2–5.

[34] Koronis, G., Chia, P. Z., Siang, J. K. K., Silva, A., Yogiaman, C., and Raghunath, N., 2019, "An Empirical Study on the Impact of Design Brief Information on the Creativity of Design Outcomes With Consideration of Gender, and Gender Diversity," ASME J. Mech. Des., **141**(7), p. 071102.

[35] Das, M., and Yang, M., 2022, "Building Confidence and Embracing Failure Through Sketching Practice," Int. J. Eng. Educ., **38**(6), pp. 1779–1790.

[36] Roldan, W., Hui, J., and Gerber, E. M., 2018, "University Makerspaces: Opportunities to Support Equitable Participation for Women in Engineering*," Int. J. Eng. Educ., **34**(2(B)), pp. 751–768.

[37] George, S. M., and Domire, Z. J., 2020, "A Six-Year Review of the Biomedical Engineering in Simulations, Imaging, and Modeling Undergraduate Research Experience," ASME J. Biomech. Eng., **142**(11), p. 111012.

[38] Reed, E. C., Kain, D., and George, S. M., 2021, "The Relationship Between Perceived Confidence, Gender, and Writing in a Biomedical Engineering Research Experience for Undergraduates Site," ASME J. Biomech. Eng., **143**(12), p. 121010.

[39] Oti, A., and Crilly, N., 2021, "Immersive 3D Sketching Tools: Implications for Visual Thinking and Communication," Comput. Graph, **94**, pp. 111–123.

[40] Song, I., Yang, J., and Shimada, K., 2014, "Development of Sketch-Based 3-D Modeling System for Rapid Generation and Evaluation of Automotive Seat Shape Using Reference Models," ASME J. Mech. Des., **136**(5), p. 051001.

[41] Bernhaupt, R., Mueller, F., Verweij, D., Andres, J., McGrenere, J., Cockburn, A., Avellino, I., et al., 2020, "Live Sketchnoting Across Platforms: Exploring the Potential and Limitations of Analogue and Digital Tools," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, Apr. 25–30.

[42] Orbay, G., and Kara, L. B., 2012, "Shape Design From Exemplar Sketches Using Graph-Based Sketch Analysis," ASME J. Mech. Des., **134**(11), p. 111002.

[43] Robertson, B. F., and Radcliffe, D. F., 2009, "Impact of CAD Tools on Creative Problem Solving in Engineering Design," Comput. Aided Des., **41**(3), pp. 136–146.

[44] Zhang, W., Ranscombe, C., Radcliffe, D., and Jackson, S., 2019, "Creation of a Framework of Design Tool Characteristics to Support Evaluation and Selection of Visualisation Tools," Proc. Des. Soc. Int. Conf. Eng. Des., **1**(1), pp. 1115–1124.

[45] Elsen, C., Demaret, J.-N., Yang, M. C., and Leclercq, P., 2012, "Sketch-Based Interfaces for Modeling and Users' Needs: Redefining Connections," Artif. Intell. Eng. Des. Anal. Manuf., **26**(3), pp. 281–301.

[46] Macomber, B., and Yang, M., 2011, "The Role of Sketch Finish and Style in User Responses to Early Stage Design Concepts," Volume 9: 23rd International Conference on Design Theory and Methodology; 16th Design for Manufacturing and the Life Cycle Conference, Washington, DC, Aug. 28–31, pp. 567–576.

[47] Li, W. W., Hilton, E., Hammond, T., and Linsey, J. S., 2016, "Persketchtivity: An Intelligent Pen-Based Online Education Platform for Sketching Instruction," Electronic Visualisation and the Arts, London, UK, July 2016, pp. 133–141.
[48] Hilton, E., Li, W., Newton, S. H., Alemdar, M., Pucha, R., and Linsey, J., 2016, "The Development and Effects of Teaching Perspective Free-Hand Sketching in Engineering Design," Volume 3: 18th International Conference on Advanced Vehicle Technologies; 13th International Conference on Design Education; 9th Frontiers in Biomedical Devices, Charlotte, NC, Aug. 21–24.
[49] Weaver, M. B., Ray, S., Hilton, E. C., Dorozhkin, D., Douglas, K., Hammond, T., and Linsey, J., 2022, "Improving Engineering Sketching Education Through Perspective Techniques and an AI-Based Tutoring Platform," Int. J. Eng. Educ., **38**(6), p. 2013612.
[50] Lau, K., Oehlberg, L., and Agogino, A., 2009, "Sketching in Design Journals: An Analysis of Visual Representations in the Product Design Process," Eng. Des. Graph. J., **73**(3), pp. 23–28.
[51] Evans, M., and Aldoy, N., 2016, "Digital Design Sketching Using the Tablet PC," Des. J., **19**(5), pp. 1–22.
[52] Badam, S. K., Chandrasegaran, S., Elmqvist, N., and Ramani, K., 2014, "Tracing and Sketching Performance Using Blunt-Tipped Styli on Direct-Touch Tablets," Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces—AVI'14, Como, Italy, May 27–30, pp. 193–200.
[53] Peng, A., Menold, J., and Miller, S. R., 2020, "Does It Translate? A Case Study of Conceptual Design Outcomes With U.S. and Moroccan Students," Vol. 3: 17th International Conference on Design Education (DEC), Virtual, Online, August 2020.
[54] Larkin, J. H., and Simon, H. A., 1987, "Why a Diagram Is (Sometimes) Worth Ten Thousand Words," Cogn. Sci., **11**(1), pp. 65–100.
[55] Linsey, J. S., Tseng, I., Fu, K., Cagan, J., Wood, K. L., and Schunn, C., 2010, "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," ASME J. Mech. Des., **132**(4), p. 041003.
[56] Zheng, X., and Miller, S. R., 2021, "Out in the Field Versus Inside in the Lab: A Comparison of Design Professionals' Concept Screening Practices," ASME J. Mech. Des., **143**(4), p. 041401.
[57] Toh, C. A., and Miller, S. R., 2013, "Visual Inspection or Product Dissection? The Impact of Designer-Product Interactions on Engineering Design Creativity," Volume 5: 25th International Conference on Design Theory and Methodology; ASME 2013 Power Transmission and Gearing Conference, Portland, OR, Aug. 4–7.
[58] Thinkspace, 2004, "Thinkspace: PeanutSheller," http://www.thinkcycle.org/tc-space/tspace?tspace_id 41963
[59] Tsenn, J., Atilola, O., McAdams, D. A., and Linsey, J. S., 2014, "The Effects of Time and Incubation on Design Concept Generation," Des. Stud., **35**(5), pp. 500–526.
[60] Guilford, J. P., 1979, "Some Incubated Thoughts on Incubation," J. Creat. Behav., **13**(1), pp. 1–8.
[61] Landis, J. R., and Koch, G. G., 1977, "The Measurement of Observer Agreement for Categorical Data," Biometrics, **33**(1), pp. 159–174.
[62] Freelon, D. G., 2013, "ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service," Int. J. Internet Sci., **1**(8), pp. 10–16.
[63] Freelon, D. G., 2010, "ReCal: Intercoder Reliability Calculation as a Web Service," Int. J. Internet Sci., **1**(5), pp. 20–33.
[64] Xu, K., Wendell, D., and Walsh, A. S., 2017, "Getting to Gender Parity in a Top-Tier Mechanical Engineering Department: A Case Study," 2017 ASEE Annual Conference & Exposition, Columbus, OH, June 25–28.