Assessing Early Stage Design Sketches and Reflections on Prototyping

Assessing Early Stage Design Sketches and Reflections on Prototyping

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

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

Designers routinely create informal "thinking" sketches to explore a design space, "talking" sketches to communicate design ideas during the early phases of the design process, and "learning" prototypes to test potential concepts. This study presents two new tools to assess novice designers' sketch attributes and prototyping reflections in the context of an introductory design course. First, it proposes a rubric for assessing the quality of early-stage design sketches including line smoothness, proportion, and understandability. Of particular note is the contribution of assessing understandability as a metric for sketches as communication tools. This study also presents a tool to capture designer reflections after each iteration of a prototype. Not only does this record what is learned about a design but also designers' personal and emotional reactions to the process. Sketching-related results show a positive correlation between sketch quality and understandability, indicating the importance of sketch quality especially when designers use sketches to communicate. Results also indicate that early-stage sketch quantity, but not quality, is linked with design outcomes. This study also finds a link between the frequency of sketching and higher maximum sketch quality scores (i.e., at least one highly rated sketch) as well as a correlation between individuals' maximum sketch quality scores and overall design outcomes. Preliminary results around prototyping indicate that reflection on both the technical and emotional aspects of prototyping may be a worthwhile area of further study. Finally, several results point to novice designers' lack of consistent focus on users in their prototyping reflections and presentations.
[DOI: 10.1115/1.4053463]

Keywords: creativity and concept generation, design evaluation, design methodology, design process, design representation, design theory, design theory and methodology, product design

1 Introduction

1.1 Motivation

1.1.1 Sketching. Exploratory sketches and prototypes are a hallmark of the early stages of the design process, and several of their aspects have been shown to be linked with design outcomes [1–3]. Many efforts have been made to evaluate sketch quality, especially for more refined sketches at later stages of design [4–9]. However, the value of specific characteristics of early design phase drawings can be difficult to assess. Of particular interest are informal "thinking" drawings that help designers clarify their own design concepts and "talking" drawings that help designers communicate their ideas, as defined by Ferguson [10]. The difficulty of assessing these types of sketches is in part due to their rough, often ambiguous nature that makes them difficult to analyze in the same way as more formal graphical representations like computer-aided design (CAD) drawings [11]. Since these thinking and talking sketches are often used as communication tools in the engineering design process, it is important to consider what sketch characteristics make them more effective for communication. To address this, this article presents a study that explores three aspects of early-stage sketches. This includes their mechanical aspects, in this case, line quality/smoothness [4,5,9], their overall proportionality/accuracy [4,5,9], and their ability to communicate and be understood by others [10]. The line smoothness and proportionality measures are used as a metric of the overall quality of the drawing, while understandability is used to evaluate the effectiveness of the sketch as a communication tool.

1.1.2 Prototyping. Prototyping is another essential component of early-stage design and is often thought about in terms of the artifact being designed. Perhaps equally important in the early stages of design is the information that the designer learns from building and testing a prototype. Much research has been conducted on different frameworks for prototyping [12,13]. However, areas that have garnered less attention are the personal impacts of the prototype on the designer, such as what they learn or what their personal reflections are on a prototype. Foundational work by Schön has demonstrated the importance of reflective practice for designers [14] as well as in engineering education for novices [15]. Reflective practice is a natural extension of iteration and is a key to synthesizing new information about a given design problem. Prior research shows that senior students have a stronger grasp on reflective practices than first-year students do, so it is crucial to help build these practices early for novice designers [16]. Prior work also shows that novice designers often do not realize the ways in which they are using prototypes until prompted to reflect on them [17].

Discussions around the psychological experience of prototyping, especially for practitioners, show that different types of prototyping experiences can lead to different outcomes [18]. It is important to evaluate this for novice designers as well and understand how to produce the kinds of prototypes that have both positive psychological outcomes and positive design outcomes, especially as novices learn and gain experience in design. This article proposes a new framework for categorizing prototypes, which builds on existing frameworks from Nelson et al. and Lauff et al. and adds in personal reflection as a key component of the prototyping process [12,13].

Contributed by the Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received June 7, 2021; final manuscript received December 29, 2021; published online February 15, 2022. Assoc. Editor: Scarlett Miller.

Both sketching and prototyping are key parts of early-stage design and finding links between these processes, and eventual design outcomes can help build a deeper understanding of the design process. Of particular interest are novice designers' experiences with sketching and prototyping as early practices may become habits for design practitioners [17].

1.2 Research Questions

RQ1: What mechanical aspects of early-stage design sketches are linked with design outcomes for novice designers?

Early-stage design sketches are typically rough, ambiguous, and evolving, making it difficult to assess such sketches and assess any links to the resulting design outcomes. This question explores whether sketch metrics based on mechanical aspects of sketches (such as line smoothness and proportion/accuracy) are linked with design outcomes. We hypothesize that sketch quality and design outcomes will be linked since designers who more effectively use sketches to clarify and communicate their ideas may also develop better designs.

RQ2: Are higher quality early-stage sketches more understandable to outside viewers?

Early-stage design sketches are often used as communication tools within design teams, between designers and clients, and also for designers to better articulate their own ideas to themselves [19]. This question investigates whether such "thinking" and "talking" sketches that are low fidelity but score high in sketch quality (in this case, having smooth lines and reasonable proportion/accuracy) are also easier to understand. To assess this, we introduce a new sketch metric called "understandability" that quantifies how easy a sketch is to understand. We then compare sketch understandability and sketch quality scores to understand if higher quality sketches are more understandable.

RQ3: Does intentional reflection on the prototyping process correlate with better outcomes for novice designers?

Prototyping is used as a tool to test out concepts and determine feasible design directions. This question explores whether novice designers who reflect on their prototyping by writing down what they learned from each iteration create better designs than those who do not. We hypothesize that novice designers who reflect during the prototyping process will learn more from the iterative process and produce better outcomes.

2 Background

2.1 Role of Sketching in Design. Previous studies show that sketching and prototyping are key parts of the engineering design process and are used by novices and practitioners alike to explore and evaluate potential design directions [20]. Both sketching and prototyping are typically used iteratively and allow designers to understand and modify concepts before investing significant time and resources into a single concept. Sketching has been identified as a way to clarify concepts and communicate them to others during the ideation process [6]. There has recently been research that suggests that expertise with sketching is more valuable than delving straightaway into CAD software to improve novices' engineering design skills and avoid premature fixation on a single design [21]. Previous research on the ideation process indicates that there are correlations between the quantity of ideas sketched and eventual design outcomes [1]. Sketches considered to be higher quality have also been shown to be perceived as more creative concepts [22].

2.2 Role of Prototyping in Design. While sketches can be used to explore a breadth of product ideas, prototypes are useful to investigate individual concepts in a greater depth and understand potential pitfalls [20,23–25]. As such, there is a value in including both sketching and prototyping practices for novice designers. Previous studies have tried to codify the various roles that prototypes play in the design process and developed initial tools for tracking prototyping to create a record of a design's evolution over time [26,27]. Previous work indicates that rapid iteration through prototyping during the early stages of a design process correlates with better design outcomes [23]. In addition, the importance of having clear goals for prototypes has been identified as a best practice and frameworks to help structure the prototyping process have been developed [27,28]. This supports the idea that prototypes are an important technical tool but need to be used intentionally and that evaluating the learning outcomes from prototypes can help identify whether a prototype was successful.

2.3 Context and Value of This Study. Due to the COVID-19 pandemic and current remote setting of educational and professional environments, it is clear that remote co-design will continue to be important and may also become more commonplace. As such, it is crucial to further study and develop best practices for remote design, including sketching and prototyping, and build tools that can help designers reflect and think about their own sketches and prototypes.

This study fills the gap in the research through two main contributions. First, it presents a rubric for assessing the quality of early-stage design sketches. In particular, it establishes a new metric around understandability to assess the effectiveness of sketches as tools for communication. Second, it proposes a new method for logging prototypes during the design process that includes an open-ended reflective component. Using data from these two frameworks, this article draws conclusions about correlations between sketching and prototyping behavior of novices and eventual design outcomes.

3 Methods

3.1 Overview. In this study, 19 novice designers in an introductory project-based design class were asked to track the ideas they explored in a design notebook and also record what they learned from each prototype they created. Sketch data and reflections on prototypes were then correlated with outcomes of design.

Figure 1 is a flowchart showing the taxonomy of how the sketching and prototyping rubrics were created and used to analyze the data collected throughout the study. The first step was to look at the existing literature along with the data collected to determine what elements might be most appropriate for a rubric. Next, observations from the initial literature survey and data collected were used to identify gaps in existing rubrics and develop a new rubric. This draft rubric was tested on a subset of the data and iterated on until there were no further changes. Once the rubric was revised and finalized, the final rubric was used by independent raters to score/log all of the data. These final ratings were used as the basis for any analyses conducted.

3.2 Study Context and Data Collection. The participants in the study were novice designers enrolled in an introductory level 8-week design class for Mechanical Engineering undergraduates at a northeastern US university. This was a project-based, hands-on design-and-build course and was run in an entirely remote setting due to COVID-19 safety requirements. Participants were provided with substantial kits of materials and tools for prototyping remotely. The full list of materials provided in the kit can be found in the Appendix. Students completed two open-ended design projects in the course, working individually from their living space for both projects. The primary focus of this study was the second project, a 5-week user-centered design project. During this project, students were asked to choose a user from their living community (such as a parent or roommate) who they would be able to interact with in person so that they could test prototypes despite the COVID-19 pandemic. Only one student was living alone and thus chose a user they could interact with regularly over Zoom

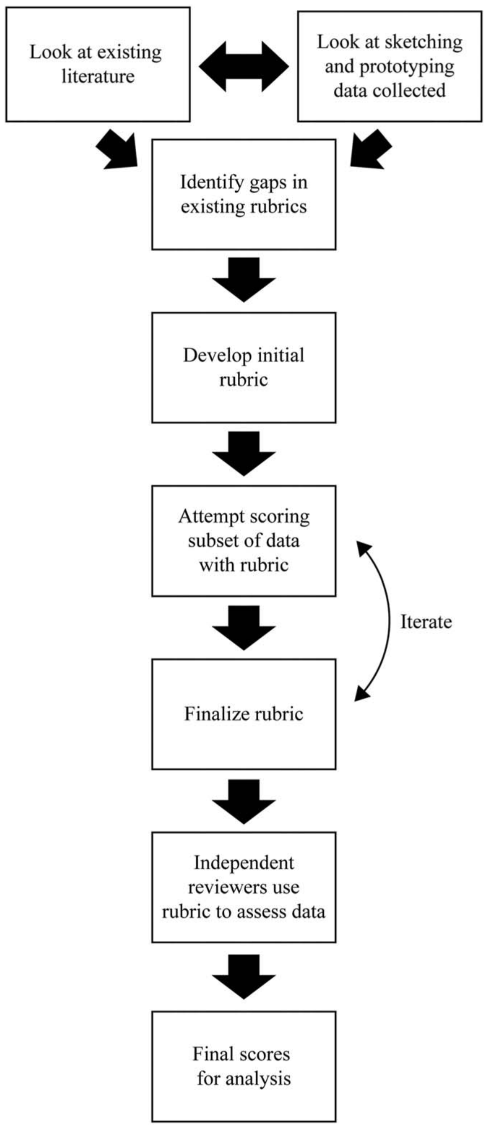


Fig. 1 Flowchart showing the process used to create the rubrics for assessing sketching and prototyping data during this study

instead. Seven men and 12 women were enrolled in the study from start to finish. The primary methods of data collection were class assignments, project assessments, grades from project reviewers, and notebook records.

3.3 Sketching. Sketching was emphasized as a key component of the design process. Students were asked to maintain a design notebook of all their project sketches and ideas that were checked by their lab instructor on a weekly basis. All sketches were done in a Rocketbook, a paper notebook that allowed participants to easily upload a digital copy of their drawings to a Google folder shared with the teaching/research team. Students were taught basic perspective drawing techniques in class and were given a short sketch practice exercise at the beginning of every class session.

The quantity of sketches made by each participant during the second design project was evaluated in two ways. The number of pages of sketches per participant was counted as was the total number of individual sketches, as each page could contain several related sketches. A single rating was given to all the sketches on each page rather than each sketch since pages with multiple related sketches tended to have consistent sketch characteristics.

To assess the quality of sketches from the class and for the second design project, a rubric was developed for this study with categories as shown in Fig. 2. Figure 2 also shows examples of sketches that fit each score for each category in the rubric. The rubric was based on existing literature [4,5,9] and an initial review of the characteristics of sketches generated in the Rocketbooks. The sketches were primarily from the brainstorming stage, often classified as "thinking" and "talking" sketches as defined by Ferguson [10]. Common themes in sketch quality metrics in the literature as applicable to this stage of the design process were consolidated. Existing rubrics included metrics for assessing the accuracy of perspective and shading [4,5,9]. Others focused on the relation between sketches, such as Goel's categorization of lateral and vertical transformations when using sketches as a way to ideate and refine concepts [8]. Another rubric was based on a computational tool for sketching and thus was able to assess metrics such as speed and stroke order coupling, which appear to be useful metrics but are not accessible when assessing hand drawings [9]. However, several metrics that were consistently used to determine sketch quality were also applicable to the type of early-stage sketches that were the focus of this study. These metrics are primarily visual and include line smoothness/confidence, as well as accuracy/proportion since accuracy and proportion were often used interchangeably in other rubrics [4,5,9]. These two metrics were adapted for the proposed sketch rubric, and their scores were summed to create a measure of overall sketch quality.

A third category of "understandability" was added to assess the ability of the sketch to serve as a communication tool. Research in the field of Artificial Intelligence has explored an adjacent concept of "sketch recognition," which is focused primarily on recognizing the dominant object in a sketch [29]. However, our new category of "understandability" explores whether a sketch can be understood in the context of the design process. We defined this as follows: "Can the rater easily understand what the sketcher tried to represent without relying on words and descriptions?" For instance, the reference image for a score of 4 in the Understandability column in Fig. 2 might be identified by sketch recognition software as a face. However, that is not very helpful in the context of the artifact being designed. It is much more important in this context to understand that the image is showing a wearable device that is emitting something. This category of "understandability" was not found in the prior literature, but it is useful to track the effectiveness of the sketch as a communication tool. This was especially salient as participants were more reliant on using sketches to communicate their design ideas to teammates and course staff in the virtual setting.

Although image contrast and clarity varied among images, as shown in Fig. 2, these were artifacts of participants' methods for uploading images digitally rather than attributes of the sketches. These aspects of the images were not considered during the sketch assessment process.

Sketches were assessed by three independent reviewers, including the authors and a graduate design student not working on this project. For this study, the raters were also the course staff for the course that was the source of data. All ratings were done after the conclusion of the course so that all sketches from the course were complete and to avoid bias from early sketch ratings influencing grading of final design solutions. Raters were given written descriptions of each category and scoring level along with example sketches that would fit each score, as shown in Fig. 2, to ensure alignment between raters.

Raters were asked to rate all sketches within a category before moving to the next category to ensure consistency between categories and avoid confounding the categories within ratings for a single sketch. This was done in the order shown in Fig. 2: line smoothness, then proportion/accuracy, and finally understandability. For the category of understandability, raters used their knowledge of the project and annotations to determine what the sketcher was trying to draw. Subsequently, they rated the understandability of the sketch by looking solely at the drawing and comparing it with

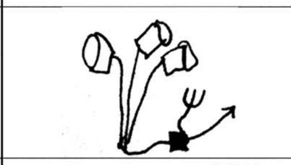
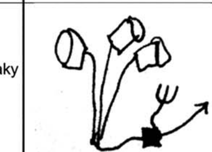
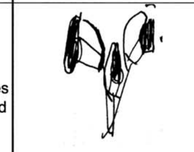
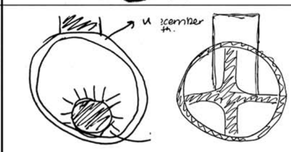
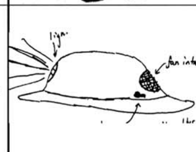
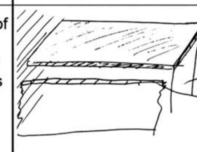
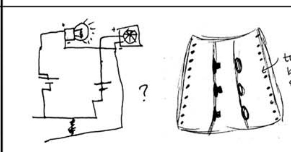
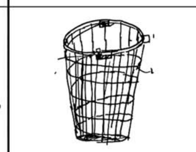
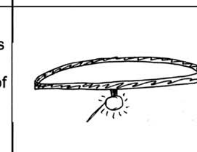
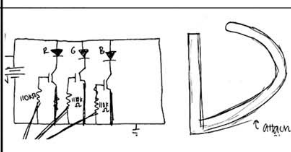
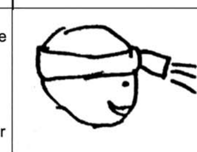
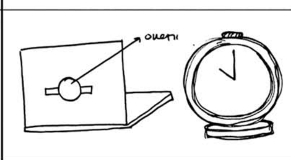
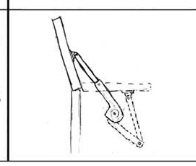
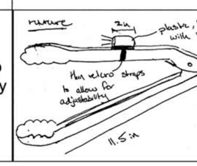
	Line Smoothness		Proportion/Accuracy		Understandability	
Definition	How smoothly (lack of waviness) each line is drawn.		How accurately each line is drawn relative to where it should be drawn.		Can the rater easily understand what the sketcher tried to represent (without relying on words and descriptions)?	
Existing Literature	[4,5,9]		[4,5,9]			
1		Lines are more often shaky than not		Sketch does not adhere to proportions; Shapes are very misaligned		Cannot even begin to guess what the object is out of context
2		Most lines are smooth; Occasional large wobbles		General contours of the sketch are correct, but there are large mistakes in overall proportional accuracy		General shapes are understood, but it is hard to guess what sketch is of
3		Lines are mostly smooth with occasional small wobbles OR Lines are hashed together, but many strokes stray from the overall contour		Overall proportions are correct, but smaller elements of the sketch are incorrect		Sketch could be representing one of a few options
4		Lines are mostly smooth with almost no wobbles OR Lines are hashed together, but a few strokes stray from the contour		Most aspects of the sketch are proprotionally accurate; Any mistakes in accuracy are minor		It is generally clear what the sketch is of, but some details are confusing
5		Lines are perfectly smooth with little exception OR Lines are hashed together, but no strokes stray from the contour		Sketch appears to the eye as perfectly proportional		Sketch is perfectly clear; Most people could guess what this sketch is of on the first try

Fig. 2 Proposed sketching rubric showing the criteria, their definitions, and any relevant existing literature that contributed to the inclusion of the criterion to the rubric along with written descriptions and examples of sketches from the sketch logbooks that would fit each scoring category. This rubric was referenced by three independent raters as they scored each page of sketches.

what they knew the sketcher was trying to represent and disregarding any text or annotation. For this study, the raters were also the course staff and thus had sufficient knowledge about each student's project that they could use to determine what the student was trying to represent in their sketch. For instance, they knew that the sketch in Fig. 2 that shows a score of 4 for understandability is meant to represent a person wearing a headlamp. However, the sketch as-is could be showing either a speaker or a light strapped to the head, so it was rated as 4 instead of 5. For further applications of this rubric, it would be important to have participants write a separate text description of the image that raters who did not have involvement in the process can use to assess understandability.

After an initial round of rating (interrater reliability from this round is discussed in the results), raters discussed discrepancies and re-rated to ensure that each rater held similar mental models of what the rubric meant and thus were assessing sketches consistently. Each participant's sketches were rated on a scale of 1–5, where 5 was the highest score for line smoothness, proportion/accuracy, and understandability. The mode of reviewer responses for each sketch was calculated (or median, if there was no mode). Line smoothness and proportion/accuracy scores were added together for a total score out of 10 to represent the "overall sketch quality."

3.4 Learning From Prototypes. To capture participants' learnings and reflections on the prototypes they created, they were assigned to individually complete an online prototype logging tool (a simple form) each time they created a new prototype. The form said, "Please fill out this form each time you create a new prototype or make a major change to an existing prototype. Note: Your prototype does not need to be in its finalized form for you to fill this out!" Participants were also given weekly reminders to continue filling out the form for the duration of the project. The form asked for the following:

- Participant name
- Date the prototype was created
- A short (one sentence) text description of the prototype
- An image of the prototype (participants could upload up to five images at once)
- A brief (one sentence) text description of what was learned from the prototype

The goal of the prototype logger was to encourage reflection and have participants articulate what was gained from each iteration. The quantity and quality of logs were not graded: students received full credit if they completed at least one log during the course.

The textual content of each prototype log was tagged according to a coding scheme to facilitate the analysis of the data. The tagging rubric was developed based on existing literature [12,13] and open coding through the initial participant responses in the prototype logger to identify major themes. Due to the individual nature of the design project and the remote setting for data collection, the prototypes were primarily "learning" prototypes, or prototypes that are used to answer questions about a potential design, as defined by Ulrich et al. [30].

The prior art shows several existing prototype planning tools. Hansen's Prototyping Planner allows designers to plan out prototyping and allow for deliberate decision-making [31]. This tool is focused primarily on planning before creating a prototype rather than reflecting on the prototype afterward. Lauff's Prototyping Canvas scaffolds prototyping planning in much more detail and

Table 1 Each theme and category of prototype assessment along with an example of text that was tagged under each category

Theme	Category	Partial example response	Existing literature
Build/test	Engineering performance (assembly or sub-assembly)	"The platform is too slippery for the iPad"	[13]
	Materials	"The wood and acrylic are strong enough"	[13]
	Users	"Allows physical user input"	[12,13]
Learning	New technical knowledge (from current iteration)	"Which libraries to install in the Arduino IDE"	[12]
Planning	Planning next design or fabrication steps (for future iteration)	"I want to try mounting it on a wall"	
Reflection	Personal reflection/emotion	"I need to get better at sewing"	
Tone	Positive	"It worked!!!"	
	Neutral	"A fan or holes is needed for ventilation"	
	Negative	"There is still much to be done but no time left"	

Note: References are included for the categories that are adapted from existing literature.

includes a reflective section on insights gained from testing with the prototype [26]. The prototype logger described in this present study was more open ended and included free-response questions rather than preselected options for participants to encourage open reflection. As a result, we focused on identifying and building on existing schemes for categorizing learning outcomes of prototypes, as presented in Table 1. Some prior coding schemas focused on specific prototyping goals such as needfinding [32] or were tailored specifically to the design outcomes of a particular experiment [33]. Others emphasized later stages of the prototyping process than the early-stage design that was the focus of this study [34]. Several exemplars were identified that were in line with the type and stage of prototyping of interest in this study. Nelson et al. used a prototyping survey based on prior work where participants self-selected goals and learning outcomes from a predetermined set of options [12]. Lauff et al. used a similar methodology as our approach by referencing existing literature and supplementing with open coding [13]. The finalized coding categories are presented in Table 1 and incorporated elements from both Nelson and Lauff's coding schemas [12,13]. Several unique categories such as planning, reflection, and tone were added based on themes that emerged from reviewing the entries in the logger and will be further discussed in the findings.

Three independent reviewers tagged each comment. The final aggregate tags for analysis were determined using the mode of the reviewers' responses (i.e., if at least two of three reviewers used a tag, it was included, and if at least two of three reviewers did not use a tag, it was not included).

3.5 Project Assessments. Final projects were presented in a video format and assessed by seven independent reviewers who are design experts (upper-level graduate students, instructors, or practitioners). Projects were given an overall score and also assessed on seven distinct metrics that were based on class deliverables (rather than being developed specifically for this study): needs, appropriateness, type of prototype, innovation, user feedback, craftsmanship, and presentation value. These categories are described in more detail in the results. All categories were assessed on a scale of 1–5, where 5 was the highest score. Final scores for the project were the mean of the scores given by all reviewers in the "overall score" category and ranged between 3.3 and 4.7 on the five-point scale. These final scores are subsequently referred to as "design outcomes." Note that these final scores do not use the average of the seven distinct metrics listed earlier and are rather only based on the results of the overall score category.

4 Results and Discussion

4.1 Sketches. The sketches assessed for this study were all "thinking sketches" and "talking sketches" from the ideation phase of the design process [10] using the rubric shown in Fig. 2. Overall sketch quality was measured on a scale of 1–10 as the sum of line smoothness and proportion/accuracy. An additional category of "understandability" was added to the sketching rubric to assess the ability of the sketch to serve as a communication tool. Spearman's Rho was used to calculate correlations between attributes of sketches and overall design outcomes due to the nonparametric nature of the data.

Krippendorff's alpha for interrater reliability was calculated using the open-source tool ReCal [35,36]. Sketches were assessed by three independent reviewers with a Krippendorff's alpha of 0.512. Landis and Koch's cutoffs can be used as a benchmarking guideline here to interpret that this alpha value signals moderate agreement between reviewers [37].

A total of 137 pages of sketches were rated for this study. The percentage of pages of sketches rated in each rating category is shown in Fig. 3(*a*). Figure 3(*b*) also shows the percentage of students with median sketch ratings at each level. Figures 3(*a*) and 3(*b*) show a relatively Gaussian distribution of scores. Median sketch ratings appear to be clustered more around 3 or 4 for all categories, whereas there are more sketches with very low (1 or 2) or very high scores (5) in the total sample. Examples of sketches for each assessment category and rating level can be found in the rubric in Fig. 2. For instance, the bottom right sketch in Fig. 2 is an example of an image that would receive a 5 score on understandability because it is very clearly a pair of tongs with a light attached to it and is easy to identify without reading the textual description.

The number of sketches made by each participant correlated with the overall outcome of their project as represented by the overall score on the project as determined by the reviewers ($r = 0.473$, $p = 0.041$). This indicates that participants who sketched more frequently had better overall scores on the project. The quantity of sketches also correlated with the perceived innovation of the project ($r = 0.741$, $p = 0.015$) as rated by reviewers. The quantity of sketches was not a factor being formally graded, so there must be an alternative explanation for this correlation. It is possible that frequent sketching leads novices to better, more novel concepts and outcomes, which would be consistent with prior work about the importance of sketch quantity [1]. It is also possible that individuals who were diligent about sketching were also diligent in other aspects of the project.

Each participant's sketches were aggregated via mean, median, and maximum to investigate if the variations in those parameters could be linked to aspects of the design process. Interestingly, participants who generated more sketches also had a higher maximum sketch quality ($r = 0.482$, $p = 0.037$), which may suggest that the more novices sketch, the more likely they are to get at least one "excellent" drawing. This finding also points to the importance of novices practicing sketching as part of the design process. Participants with a higher maximum sketch quality score also had higher scores for that design project overall ($r = 0.646$, $p = 0.003$), which is an exciting finding and should be investigated further to determine if there are causal relationships. This relationship between sketch quality and design outcomes (represented here by scores on the design project) is thematically similar to prior literature that shows that sketch quality is linked to idea creativity and warrants further study [4,22]. Surprisingly, these correlations did not persist for the median and means of participant sketch quality.

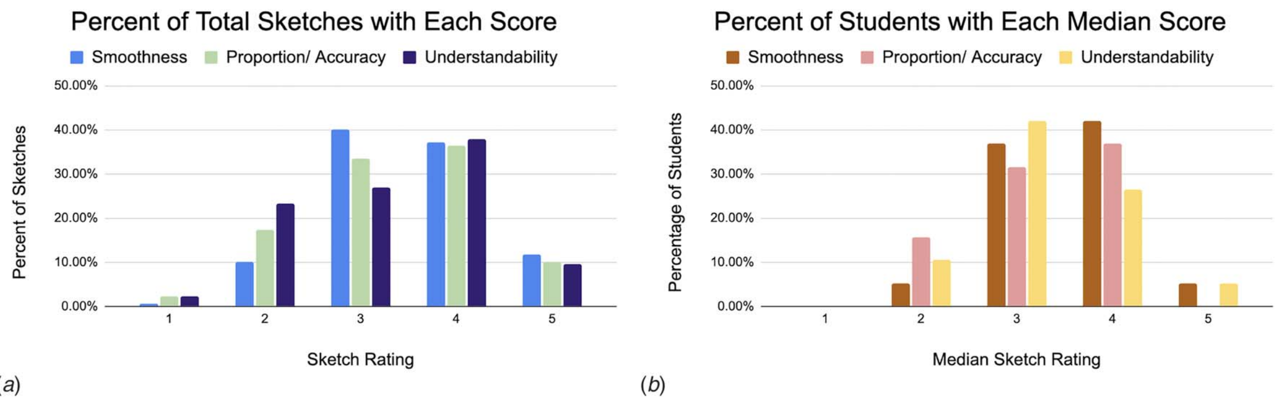


Fig. 3 (*a*) Percent of total sketches rated at each point on a scale of 1–5 (where 5 is the best score) for sketch smoothness, accuracy/proportion, and understandability and (*b*) Percent of students with median sketch ratings at each point on a scale of 1–5 (where 5 is the best score) for sketch smoothness, accuracy/proportion, and understandability

As shown in Fig. 3(*b*), most participants had similar median sketch ratings around 3 or 4 (with a similar trend for mean values), so it is possible that this lack of variation of median and mean resulted in a more homogeneous set of scores that did not allow for meaningful correlation with overall outcomes.

Mean sketch quality and mean understandability were strongly linked ($r = 0.577$, $p = 0.010$), indicating that higher quality sketches were perceived as also being more understandable. Furthermore, mean understandability had strong links with both the mean line smoothness ($r = 0.536$, $p = 0.002$) and mean sketch proportion/accuracy ($r = 0.542$, $p = 0.002$) values independently as well. This indicates that sketches with high understandability also tended to have high proportion/accuracy and high smoothness scores. This implies that if the sketch is easy to understand, it may be perceived as a higher quality sketch in other categories (and vice versa). A similar correlation was found for median sketch quality and median understandability ($r = 0.516$, $p = 0.024$) but not for maximum sketch quality and maximum understandability. Figure 4 shows examples of sketches where understandability scores were not aligned with the other ratings, showing that the understandability metric was distinct and provided additional information about the sketch rather than being confounded with sketch quality. Reviewers were not simply assessing whether a sketch was good, but instead were judging the sketch's standalone ability to communicate the concept effectively.

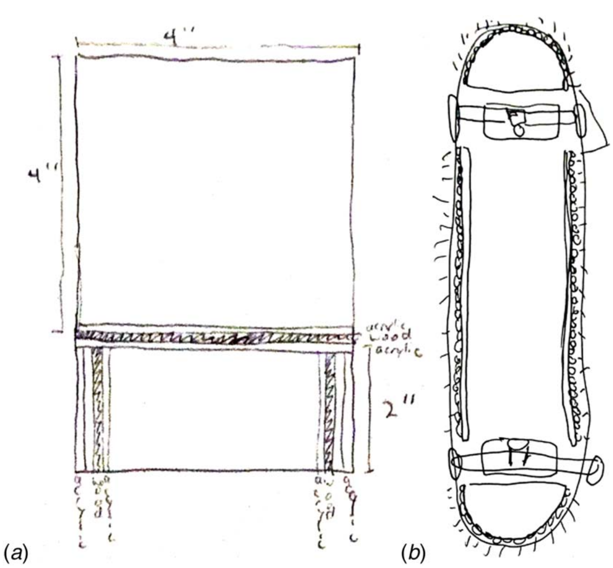


Fig. 4 (*a*) A sketch representing the side view of a phone stand that received high smoothness and proportion/accuracy scores (4 and 5, respectively) but a low understandability score (2) and (*b*) A sketch representing a light-up skateboard that received low smoothness and proportion/accuracy scores (2 and 3, respectively) but a high understandability score (5).

4.2 Reflections on Prototypes. As described in Sec. 3, the prototype logger was used by participants to record each unique prototype made and what was learned from the prototype. Seventy-three prototype logs were submitted by all participants in total. Table 2 shows example images that were submitted to the logger of a project at different stages. One prevalent theme in the reflections on each prototype that was not found in the existing research literature was the idea of planning for the next design or fabrication steps. It was observed that participants would use the open-ended textbox in the logger to not only describe what they had done but also describe what they needed to do next. Comments such as "I also want to try giving it different reflective materials that are easily removable for my next prototype" and "I need to make a stronger base or make a heavy weight" indicate that participants used this reflection space as a chance to think through and articulate next steps for their design direction.

Unexpectedly, many subjects included personal reflections or emotional comments as well, possibly due to the open-ended nature of the tool. These included observations about their own skills, frustrations, and comments that were often unrelated to the actual prototype itself. Finally, many of these comments had clear positive or negative messages, so the overall tone of each comment was tagged as positive, negative, or neutral.

Prototype logger comments were categorized by three independent raters with Fleiss' Kappa for interrater reliability of 0.633 and average pairwise percent agreement between reviewers of 82.80% [35,36]. This Kappa value and pairwise percent agreement indicate substantial agreement between reviewers [37]. Each overall theme and comment category are listed in Table 1 above along with an example response that was tagged in that category and the relevant literature used to develop the rubric. Each log was tagged for every theme in the rubric that it included, giving it a binary yes/no value for each category. Table 3 presents the final number and percentage of comments in each category.

There was a substantial difference in the quantity of entries people made in the prototyping logger, ranging from 0 to 14 per participant. In addition, the level of detail used to describe learning outcomes in the prototype logger varied widely. For instance, one participant simply wrote "Gave me a better idea of where I want to go with this" for one entry. Another provided a 200+ word explanation of specific mechanical and electronics debugging for a

Table 2 Photos that a student submitted in the prototype logger along with the submitted learning outcomes and the categories they were tagged as. On the left is a desktop sketch model to test the concept of a folding blind to block out light from the window. On the right is the final full-scale working model.

Image		
Description of what was learned from the prototype	"The foam core is hard to control when folding together, and doesn't hold its shape when folded out."	"While the frame did a good job of acting as a curtain guide, I also expected the front pieces to block light from coming in the top and bottom of the frame. However, it allowed a LOT of light through the top and bottom where the folds in the foam core allowed a lot of light to slip through the top and bottom."
Categories the comment was tagged under	Engineering performance, materials, negative tone	Engineering performance, materials, personal reflection/emotion, neutral tone

particular iteration and what was learned from each step in the debugging process. The average entry contained around 36 words.

Table 3 shows how often each category was referenced in the comments from the prototype logger. Engineering performance is referenced in the majority of the comments. Materials, planning next iterations, and personal reflections/emotions all also came up frequently. Especially interesting is the prevalence of personal reflections and emotions, as this is not typically formally tracked in the design process. For instance, comments such as "there is still much to be done but no time left" or "this thing is actually going to work and I didn't think it would" show that the prototype logger was a space where participants processed their reactions to their work in addition to articulating the technical details of what they had done.

Surprisingly, very few comments in the prototype logger referenced end users. Prior research has shown that novice designers often struggle with using prototypes to define user requirements [17]. However, we had expected that since students were required to work with users they were living with (due to the COVID-19 pandemic), they would have an easier time using prototypes for user feedback. Prior work has shown that novice designers who engage with more diverse information sources (rather than simply relying on sources such as the internet) have higher stakeholder validity scores, again emphasizing the importance of engaging with users [38,39]. It is possible that these novice designers did not leave sufficient time for user testing or felt that their prototypes needed to have a higher level of finish before testing with users or that they did not grasp the utility of getting feedback on their prototypes. This is a common misconception for novices in the design process, and efforts should be made to dispel these beliefs and emphasize the importance of testing with users regularly throughout the process [24].

Table 3 Absolute number and percentage of comments in the prototype logger corresponding to each category

Theme	Category	Total number of Comments	Percent of Comments
Build/test	Engineering performance (assembly or sub-assembly)	63	86.3
	Materials	29	39.7
	Users	11	15.1
Learning	New technical knowledge	4	5.5
Planning	Planning next design or fabrication steps	28	38.4
Reflection	Personal reflection/emotion	29	39.7
Tone	Positive	13	17.8
	Neutral	49	67.1
	Negative	11	15.1

In another unexpected finding, a minimal percentage of comments describe new technical knowledge learned in the process of creating a prototype. Nelson and Menold found a difference in teams' likelihood of reporting new technical knowledge in beta versus alpha prototypes based on their previously stated goals of learning about technology [40]. In the case of our study, prototypes were tracked throughout the entire design cycle rather than at two specific points in the process. However, this points to several avenues of interesting further work. First, the analysis of goals and their alignment with skills gained as reported in open-ended comments would be worth exploring. Second, it would be interesting to see if prototypes at different levels of fidelity (beyond alpha and beta) show different levels of technical knowledge gained.

Participants' prototype reflections showed some of their personal reactions to setbacks in addition to the technical learning on how to address issues with their prototype. As expected, the majority of comments are neutral in tone. However, the quantity of prototypes logged was positively correlated with percentage of nonneutral (positive or negative) comments ($r = 0.649$, $p = 0.003$), as those who logged prototypes more often included strong positive or negative comments rather than focusing solely on the prototype and its technical specifications. This points to a hypothesis that the open-ended setup of the prototype logger could result in increased participant reflection on prototypes.

Although there was a positive trend between the number of logs completed by each participant and overall design outcomes (scores on the project), there was not a statistically significant correlation found. The quantity of prototypes logged was not factored into student grades, but based on discussions with the participants, it is evident that more prototypes were created than were actually logged, although we do not know how many more. As a result, there are students who completed several iterations but did not fill out the logger and subsequent reflection. This imbalance presents a challenge in performing further meaningful analyses on the data. Interestingly, the number of sketches created by participants did not correlate with the number of prototypes logged, which further points to the lack of consistency with logging as a potential source of error in this study.

4.3 Assessment of Designs. The analysis of the assessment metrics used was performed using a Spearman's correlation matrix in MATLAB. The analysis reveals that scores in many of the categories are correlated with one another, which may indicate that categories are being confounded and that assessment metrics may need to be more clearly defined to be independent. The categories and their abbreviations are listed in Table 4. Figure 5 shows all the cases in which there is a statistically significant ($p \leq 0.05$) positive correlation between categories. Note that the final scores used to represent design outcomes for this study do not use the scores from these categories. Instead, they are based on the results of the overall score category, which is not represented here since it was a holistic assessment that should correlate with these categories rather than a distinct category of assessment.

Figure 5 shows r values for the categories that were statistically significantly correlated ($p \leq 0.05$). Many of the categories had strong correlations with r values of more than 0.5. This could be explained in a variety of ways. It is possible that reviewers confounded categories and tended to give participants uniformly low or high scores regardless of the specific definition of the category, especially as they were trying to rate in real-time during presentations. For instance, the novelty of the solution (Innovation) was correlated with Craftsmanship and Presentation Value, which could indicate that nicely constructed prototypes or impressive presentations were conflated with the novelty of the actual concept. Similar results have been documented between sketch quality and perceived creativity, so it is troubling to see that these results persist at the prototype level as well [4]. Innovation was also correlated with Appropriateness, although this may be more indicative of overlap in the qualities assessed by these categories since the goal of the projects was to design something for a particular user's specific need. Innovation assesses novelty, and Appropriateness assesses whether the concept met the need, so in this case, since the needs were very specific, the solutions may have been more likely to also be novel. It is also possible that participants who performed well in certain categories were more likely to perform well in other categories. For instance, it is logical to surmise that a product that "identifies a real user need" (Needs) would also likely "address the need" (Appropriateness). However, it is more surprising that both Needs and Appropriateness are correlated with Presentation Value. Again, this could be due to conflation of the categories or because participants who spent the time up front identifying and addressing a user need had a more compelling presentation to share. Type of Prototype was also correlated with every category except User Feedback. For some categories, this makes sense: an understanding of what type of model to choose may also indicate a clear understanding of the need and how to address it since different types of prototypes are used to address different design questions (Needs and Appropriateness) [27]. Similarly, clarity regarding the type of model could be easily tied to the understanding of how to effectively fabricate the model (Craftsmanship). For other categories, this is more surprising: does the correlation between the type of prototype and presentation value imply that certain types of prototypes are better for communicating design solutions? In either case, it is possible that this rubric is overassessing certain elements of participant designs and should be re-evaluated and modified.

The statistical range of all the scores in each category was also calculated to determine if there were any categories with notable discrepancies among the participants. These values range from 1.29 to 2.57 with a median range of 1.57 among the categories. User Feedback has the highest range, which indicates that there was the most variation in the extent to which participants communicated incorporating User Feedback into their designs. User Feedback is also the only category that had no correlation with other categories, perhaps because it is most easily defined in a binary fashion. This is in strong contrast to the Needs category, which had high correlation coefficients with three other categories. In particular, it is surprising that Needs and User Feedback were not correlated since the Needs category is specifically about user needs. These results may indicate a lack of emphasis on the user and mirror initial results from the prototyping logger, where very few participants referred to users in the prototyping logs. Although it

Table 4 Each category of assessment for the final design project and its shorthand label along with the framing used by the reviewers to understand the meaning of the category

Category	Abbreviation	Explanation of category given to reviewers
Needs	N	Did the student identify a real user need?
Appropriateness	A	Does the concept address the need?
Type of Prototype	TP	Is the type of model (looks-like, works-like, etc.) appropriate for addressing the key unknowns?
Innovation	I	Is the solution novel?
User Feedback	UF	Did the student collect user feedback and incorporate it into the current design, or vision for a future design?
Craftsmanship	C	Is the project well designed and fabricated?
Presentation Value	PV	Is the product story communicated in an effective way?

	N	A	TP	I	UF	C	PV
N		0.84	0.70				0.78
A			0.75	0.58			0.82
TP				0.60		0.58	0.73
I						0.49	0.54
UF							
C							
PV							

Fig. 5 Grid of relationships between design assessment metrics. Cases in which there is a statistically significant positive correlation ($p \leq 0.05$) between categories are white and list the r correlation coefficients in the grid square. Cases where there is no statistically significant positive correlation ($p \leq 0.05$) between categories are light gray. Abbreviations for each category are listed in Table 4.

is possible that participants had more user feedback than they presented or communicated in their prototype logs, this clearly points to a gap in novice designers' emphasis on the value of user feedback, as has been noticed in other previous works [17,38].

4.4 Limitations. This was a study with a small number of participants and is meant to be used as a preliminary step toward other studies. A limitation of this study as it pertains to prototype logging was that several participants did not log all of the prototypes they created. In addition, the study was run in the context of an academic course, so it was not possible to have a control group for experiments.

Another limitation of the study is in the assessment of understandability ratings. For these ratings, reviewers were asked not to look at any text or annotations in the images. However, it is possible that reviewers were unconsciously influenced by the text and annotations present on the drawings. Furthermore, in design communication, sketches are often used as communication tools in combination with written and verbal communication rather than as standalone artifacts. For the purpose of standardization, this study examined understandability solely for standalone sketches. In practice, it is possible that a lower understandability standalone sketch could be combined with annotation or verbal communication that would still make it an effective communication tool.

5 Conclusions and Future Work

This study investigated several aspects of novice designers' early-stage design processes in the context of an introductory design course.

5.1 Contributions to the Field. This study has several key contributions to the field of Design Theory and Methodology. The study developed a new rubric to assess the sketch quality of thinking and talking sketches through measures of line smoothness, proportion/accuracy, and understandability. In particular, this study establishes "understandability" as a new metric to assess the ability of a standalone sketch to communicate the key aspects of a design. The study also developed a novel tool to capture open-ended reflective comments from designers during the iterative prototyping process. This tool allows for technical and personal reflection and can be used to track iteration along with the personal growth during the process. Both the sketching rubric and prototyping logger are new tools that can be used to assess critical aspects of sketches and prototypes made during early-stage design. Results from both components of the study are presented in the context of each research question below.

5.2 Conclusions

RQ1: What mechanical aspects of early-stage design sketches are linked with design outcomes for novice designers?

This study has proposed a framework for assessing the quality of "thinking" and "talking" sketches during the ideation phase of the design process. The proposed sketch quality framework includes categories for line smoothness and proportion/accuracy. We showed that in this case, the quantity of sketches correlates with design outcomes (represented here by overall scores on their design project) for novices, but that the quality of sketches does not have a strong correlation with design outcomes. An interesting finding was that participants who sketched more had higher maximum sketch quality scores (i.e., at least one excellent sketch) and that maximum sketch quality scores also correlated with overall design outcomes.

RQ2: Are higher quality early-stage sketches more understandable to outside viewers?

The study introduced a new measure of "understandability" as a criterion for evaluating sketches as it is a key for using sketches as a communication tool. In this study, sketch quality, line smoothness, and proportion/accuracy were all strongly correlated with sketch understandability. This suggests that opportunities for novices to get practice and training in these sketching techniques may help them in creating sketches that are more understandable and effective as a communication tool in the engineering design process. This finding may also have implications for studies around design teams as it is possible that higher quality sketches would be better communication tools between team members and could lead to better designs.

RQ3: Does intentional reflection on the prototyping process correlate with better outcomes for novice designers?

Finally, the study developed a tool for designers to log prototypes during the design process, including open-ended space to describe what was learned from the prototype. This study proposes a rubric for categorizing these comments that include new categories for assessing tone, personal reflections/emotional comments, and outlining of tasks for future iterations. The prototype logging process led to comments about the technical quality of prototypes and personal reflections. These personal aspects are of particular relevance in this current pandemic time when many designers work individually and often in isolated settings with reduced interactions with colleagues, but may also apply to virtual teams that are likely to be an important work mode in the future. However, due to the fact that not all participants submitted a log for each prototype, we are unable to draw formal conclusions on whether building in reflection during the prototyping process correlates with better outcomes for novice designers.

5.3 Future Work. Many correlational relationships were identified in this study that could be starting points for further experiments to determine causal relationships. Future work would involve running these experiments in a controlled setting with novices outside of a class setting. In addition, it would be interesting to see if similar results hold for practitioners, or if these observations are unique to the novice experience in design. This work explores new, but separate, tracking tools for sketching and prototyping during a single design project. Future work should explore a common rubric that can assess both sketching and prototyping since both are often used at the same design stages. Integrating both sketching and prototyping into a single rubric would allow for comparison between the different methods of design representation and could provide additional insights on design outcomes.

Researchers should continue developing tools that allow easy tracking of prototypes for novices and practitioners alike and should consider implementing open-ended reflective questions as part of these tracking tools [12]. This is especially important as engineering design portfolios are gaining traction in industry and academia. These portfolios often have a focus on showing design processes and iterations rather than simply end products, so tools to facilitate tracking of design decisions during prototyping are especially useful.

Future studies should also examine differences in sketch quality and quantity between hand sketches and electronic sketches on tablets as tablets are gaining traction among designers. In particular, it would be interesting to note if the sketch quality metrics developed for this study are appropriate for tablet sketches as well. Similarly, it would be compelling to explore if the correlations between sketch quantity and design outcomes persist when drawings are done on tablets.

Despite the emphasis on the importance of the user during the design process, results from the prototype logger and design assessments show that novice designers are not consistently emphasizing users in the design process. This is evident in both their prototyping reflections and presentations. This is an area of concern and should be of note to those training novice designers.

Acknowledgment

We thank the students in the introductory design course for their participation in the course and willingness to provide feedback.

Additional thanks to Jana Saadi for serving as a rater for the sketches and prototype logger and Qifang Bao for advice on data analysis. This project was funded by an MIT Abdul Latif Jameel World Education Lab (J-WEL) Grant in Higher Education Innovation. Any opinions, findings, and conclusions/recommendations expressed here are those of the authors and do not necessarily reflect the views of the funders.

Appendix: Remote Learning Kit Lists

General Items Kit	Design Project 1 Kit	Design Project 2 Kit
Olfa Cutter Heavy Duty L-1 Model 5003	Assorted rubber bands	LED light strips (16.4 ft)
Zipper bag	Paper clips (100 count)	Acrylic sheet 5 in. × 7 in. (5 sheets)
15 in. aluminum ruler	Balloons (10)	Balsa wood 150 × 100 mm (six sheets)
Prismacolor PC 935 black pencil	Brads (40)	Arduino starter kit
Sharpener	Popsicle sticks (50)	Mini LED flashlight
Eraser	18 gauge craft wire	LED light bulb (pack of 4)
Cutting mats	Assorted springs	Light bulb socket
20 Watt Glue Gun with 30 glue sticks	9 in. aluminum pie plate (4)	Polyform Sculpey III polymer clay (4 blocks)
Scotch magic tape	Gobstopper 5 oz	1 in. wide elastic 11 yards
Double-sided foam tape (1in × 3yrd)	Mike & Ike 5 oz	1 yard black fabric
Duct tape 2 in. × 30 yd	Skittles 2.17 oz (3 packs)	Small sewing kit
Loctite Super Glue 20-g bottle	M&M's 3.14 oz (3 packs)	AA batteries (pack of 8)
Scissors	Disposable plastic cups (12)	AA battery holder
Twisted nylon string 500 ft	Edible glitter	AAA batteries
Compass		Alligator clips (10 wires)
Measuring tape 12 ft		Acrylic paint (6 colors)
Needle nose plyer		Paint brushes
Printer paper (400 sheets)		Laser pointer
Safety glasses		
Rocketbook		
Elmer's rubber cement		
Pentel sign pen, fiber-tipped, black ink		
Sharpies (fine point)		
Sharpie ultra-fine tip felt tip pen		
Cardstock (40 sheets)		
Post it notes (2 pads)		
Colored Sharpie fine tip (12 colors)		
Sandpaper assorted grit (6 sheets)		
18-in-1 multi-tool snowflake		
Hammer multi-tool		
Mini Hacksaw		
Foam core 24 in. × 18 in. (3 sheets)		

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.

References

[1] 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.
[2] Yang, M. C., 2009, "Observations on Concept Generation and Sketching in Engineering Design," Res. Eng. Des., **20**(1), pp. 1–11.
[3] 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.
[4] 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.
[5] 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, Georgia Institute of Technology, Atlanta, GA, Nov. 2–4.
[6] Rodgers, P. A., Green, G., and McGown, A., 2000, "Using Concept Sketches to Track Design Progress," Des. Stud., **21**(5), pp. 451–464.
[7] Sevier, D. C., Jablokow, K., McKilligan, S., Daly, S. R., Baker, I. N., and Silk, E. M., 2017, "Towards the Development of an Elaboration Metric for Concept Sketches," ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Cleveland, OH, Aug. 6–9, p. V003T04A006.
[8] Goel, V., 1995, *Sketches of Thought*, MIT Press, Cambridge, MA.
[9] 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. Tiis., **8**(3), p. 19.
[10] Ferguson, E. S., 1994, *Engineering and the Mind's Eye*, MIT Press, Cambridge, MA.
[11] Hilton, E. C., Gamble, T., Li, W., Hammond, T., and Linsey, J. S., 2018, "Back to Basics: Sketching, Not CAD, Is the Key to Improving Essential Engineering Design Skills," Volume 7 30th International Conference on Design Theory and Methodology, p. V007T06A051.
[12] Nelson, J., Berlin, A., and Menold, J., 2019, "ARCHIE: An Automated Data Collection Method for Physical Prototyping Efforts in Authentic Design Situations," p. V007T06A044.
[13] Lauff, C., Kotys-Schwartz, D., and Rentschler, M. E., 2017, "Perceptions of Prototypes: Pilot Study Comparing Students and Professionals," p. V003T04A011.
[14] Schön, D., 1983, *The Reflective Practitioner: How Professionals Think in Action*, Temple Smith, London.
[15] Turns, J., Sattler, B., Yasuhara, K., Borgford-Parnell, J., and Atman, C., 2014, "Integrating Reflection into Engineering Education," pp. 24.776.1–24.776.16.
[16] Adams, R. S., Turns, J., and Atman, C. J., 2003, "Educating Effective Engineering Designers: The Role of Reflective Practice," Des. Stud., **24**(3), pp. 275–294.
[17] Deininger, M., Daly, S. R., Sienko, K. H., and Lee, J. C., 2017, "Novice Designers' Use of Prototypes in Engineering Design," Des. Stud., **51**, pp. 25–65.
[18] Gerber, E., and Carroll, M., 2012, "The Psychological Experience of Prototyping," Des. Stud., **33**(1), pp. 64–84.
[19] Bamberger, J., and Schön, D. A., 1983, "Learning as Reflective Conversation With Materials: Notes From Work in Progress," Art Educ., **36**(2), p. 68.

[20] Bao, Q., Faas, D., and Yang, M., 2018, "Interplay of Sketching & Prototyping in Early Stage Product Design," Int. J. Des. Creativity Innov., **4**(3), pp. 146–168.
[21] 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.
[22] Kwon, J., and Kudrowitz, B., 2019, "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. 18–21, American Society of Mechanical Engineers, vol. 59278, p. V007T06A005.
[23] Neeley, W. L., Lim, K., Zhu, A., and Yang, M. C., 2013, "Building Fast to Think Faster: Exploiting Rapid Prototyping to Accelerate Ideation During Early Stage Design," American Society of Mechanical Engineers, p. V005T06A022.
[24] Häggman, A., Honda, T., and Yang, M. C., 2013, "The Influence of Timing in Exploratory Prototyping and Other Activities in Design Projects," p. V005T06A023.
[25] Yang, M. C., 2005, "A Study of Prototypes, Design Activity, and Design Outcome," Des. Stud., **26**(6), pp. 649–669.
[26] Lauff, C., Menold, J., and Wood, K. L., 2019, "Prototyping Canvas: Design Tool for Planning Purposeful Prototypes," Proc. Des. Soc. Int. Conf. Eng. Des., **1**(1), pp. 1563–1572.
[27] Menold, J., Jablokow, K., and Simpson, T., 2017, "Prototype for X (PFX): A Holistic Framework for Structuring Prototyping Methods to Support Engineering Design," Des. Stud., **50**, pp. 70–112.
[28] Menold, J., Simpson, T. W., and Jablokow, K. W., 2016, "The Prototype for X (PFX) Framework: Assessing the Impact of PFX on Desirability, Feasibility, and Viability of End Designs," p. V007T06A040.
[29] Li, Y., and Xu, W., 2021, "Using CycleGAN to Achieve the Sketch Recognition Process of Sketch-Based Modeling," pp. 26–34.
[30] Ulrich, K. T., Eppinger, S. D., and Yang, M. C., 2020, *Product Design and Development*, McGraw-Hill Education, New York.
[31] Hansen, C. A., Jensen, L. S., Özkil, A. G., and Pacheco, N. M. M., 2020, "Fostering Prototyping Mindsets in Novice Designers With the Prototyping Planner," pp. 1725–1734.
[32] Tiong, E., Seow, O., Teo, K., Silva, A., Wood, K. L., Jensen, D. D., and Yang, M. C., 2018, "The Economies and Dimensionality of Prototyping: Value, Time, Cost and Fidelity," p. V007T06A045.
[33] Menold, J., Berdanier, C., McComb, C., Hocker, E., and Gardner, L., 2018, "'Thus, I Had to Go With What I Had': A Multiple Methods Exploration of Novice Designers' Articulation of Prototyping Decisions," p. V007T06A047.
[34] Lauff, C., Kotys-Schwartz, D., and Rentschler, M. E., 2018, "What Is a Prototype? What Are the Roles of Prototypes in Companies?," ASME J. Mech. Des., **140**(6), p. 061102.
[35] Freelon, D. G., 2013, "ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service," Int. J.f Intern. Sci., **1**(8), pp. 10–16.
[36] Freelon, D. G., 2010, "ReCal: Intercoder Reliability Calculation as a Web Service."
[37] Landis, J. R., and Koch, G. G., 1977, "The Measurement of Observer Agreement for Categorical Data," Biometrics, **33**(1), pp. 159–174.
[38] Mohedas, I., Daly, S. R., and Sienko, K. H., 2015, "Requirements Development: Approaches and Behaviors of Novice Designers," ASME J. Mech. Des., **137**(7), p. 071407.
[39] Génova, G., Fuentes, J. M., Llorens, J., Hurtado, O., and Moreno, V., 2013, "A Framework to Measure and Improve the Quality of Textual Requirements," Requir. Eng., **18**(1), pp. 25–41.
[40] Nelson, J., and Menold, J., 2020, "Opening the Black Box: Developing Metrics to Assess the Cognitive Processes of Prototyping," Des. Stud., **70**, p. 100964.