Assessing Early Stage Design Sketches and Reflections on Prototyping

Proceedings of the ASME 2021
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC-CIE2021
August 17-19, 2021, Virtual, Online

DETC2021-66748

ASSESSING EARLY STAGE DESIGN SKETCHES AND REFLECTIONS ON PROTOTYPING

Madhurima Das and Maria C. Yang

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT
Designers routinely create informal "thinking" sketches to explore a design space and "talking" sketches to communicate design ideas during the early phases of the design process. This study proposes a rubric for assessing the quality of novice designers' early stage design sketches including line smoothness, proportion, and understandability. The study finds a positive correlation between sketch quality and understandability, which indicates the importance of sketch quality when using sketches as a communication tool. Results indicate that early stage sketch quantity is linked with design outcomes, though sketch quality does not have a strong correlation with design outcomes. The study also finds a link between frequency of sketching and having higher maximum sketch quality scores (i.e. at least one excellent sketch) as well as a correlation between individuals' maximum sketch quality scores and their overall design outcomes. This study presents a new tool to capture what is learned by the designer after each iteration of a prototype. Preliminary results indicate that reflection on both the technical and emotional aspects of prototyping may be valuable and should be an area of further study. Finally, several results point to novice designers' lack of consistent focus on users in their prototyping reflections and presentations.

1. INTRODUCTION
1.1 Motivation
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]. Several 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, particularly informal "thinking" and "talking" drawings, as defined by Ferguson, can be difficult to assess [10]. This is in part due to the rough, often ambiguous nature which makes them difficult to analyze in the same way more formal graphical representations like CAD drawings can be analyzed [11]. Since these thinking sketches are often used as a communication tool in the engineering design process, it is possible that higher quality sketches are also easier to understand. To address this, this paper presents a study that explores three aspects of early stage sketches. This includes their mechanical aspects, in this case, line quality [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 the understandability is used to evaluate the effectiveness of the sketch as a communication tool.

Prototyping is another essential component of early stage design and is often thought about in terms of the artifact being designed, but 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, one area that has garnered less attention are the personal impacts of the prototype on the designer, such as what they learn or how they might reflect personally on a prototype. Discussions around the psychological experience of prototyping, especially for practitioners, shows that different types of prototyping experiences can lead to different outcomes [14]. 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. This paper proposes a new framework for categorizing prototypes that builds on existing frameworks from Nelson and Lauff but adds personal reflection as a key component [12,13].

Both sketching and prototyping are key parts of early stage design and finding links between these processes and eventual design outcomes is of much value to the design community. Of particular interest is novice designers' experiences with sketching and prototyping as early practices may become habits for design practitioners.

Research Questions:
RQ 1: *What aspects of early stage design sketches are linked with positive design outcomes for novice designers?*
Early stage design sketches are typically rough, ambiguous, and evolving, making it difficult to assess such sketches to predict the potential success of the resulting design. This question explores whether a simpler sketch metric based the mechanical aspects of sketches can be linked with design outcome.

RQ 2: *Are higher quality sketches easier to understand?*
Early stage design sketches are often used as communication tools between teams or between designers and clients. This question investigates whether "thinking" and "talking" sketches that are low fidelity but scored as having smooth lines and reasonable proportion/accuracy are also easier to understand.

RQ 3: *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 thoughtfully reflect on their prototyping create better designs than those who don't.

1.2 Background

1.2.1 Role of Sketching in Design
Prior 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 [15]. Both sketching and prototyping are typically used iteratively and allow designers to understand and modify concepts before investing significant time and resources to 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 Computer-Aided Design (CAD) software in order to improve novices' engineering design skills and avoid premature fixation on a single design [11,16]. Prior research on the ideation process indicates that there are correlations between 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 [17].

1.2.2 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 greater depth and understand potential pitfalls [15,18–20]. As such, there is value in including both sketching and prototyping practices for novice designers. Prior studies have also tried to codify the various roles that prototypes play in the design process and developed tools for using prototyping information to create a record of a design's evolution over time [21,22]. Prior work indicates that rapid iteration through prototyping during the early stages of a design process correlates with better design outcomes [18]. Additionally, 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 [22,23]. 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 or not a prototype was successful.

1.2.3 Context and Value of this Study
Due to the 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, including tools that can help designers reflect and think about their own sketches and prototypes.

This study fills the gap in the research by creating a rubric for assessing 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. Additionally, the paper proposes a new method for logging prototypes during the design process that includes an open-ended reflective component. The paper draws conclusions about correlations between sketching and prototyping behavior of novices and eventual design outcomes.

2. METHODS

Overview: In this study, 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.

2.1 Study Context and Data Collection

The participants in the study were novice designers enrolled in an introductory level eight-week design class for Mechanical Engineering undergraduates at a northeastern US university. The course was a project-based, hands-on design-and-build course and was run in an entirely remote setting due to COVID-19 requirements. Participants were provided with substantial kits of materials and tools for prototyping remotely. Students completed two open-ended design projects in the course, working primarily on their own from their home or dorm. The primary focus of this study was one of the projects, a 5-week user-centered design project. Seven men and twelve 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.

2.2 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 which 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 was counted as was the total number of individual sketches, as each page could contain several sketches.

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 Figure 1. Figure 1 also shows examples of sketches that fit each score for each category in the rubric. The rubric was based on existing literature 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 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 were 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 [4,5,9]. Another used computational tools 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 thinking sketches that were the focus of this study. These metrics are primarily visual and include line smoothness/confidence, as well as accuracy/proportion as 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 was added to assess sketch understandability and will be discussed further in the findings.

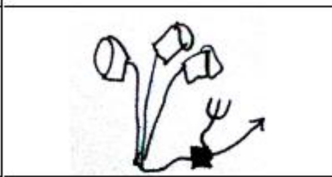
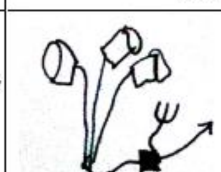
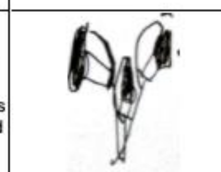
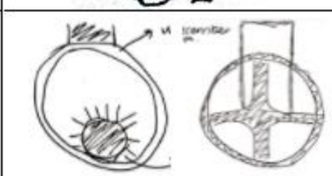
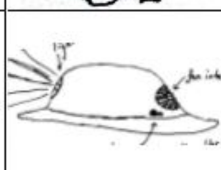
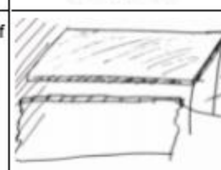
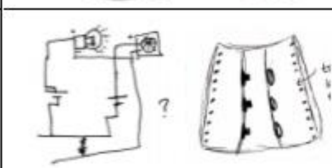
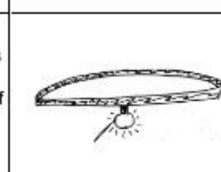
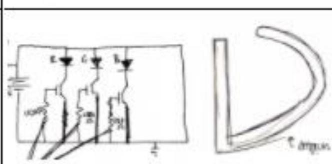
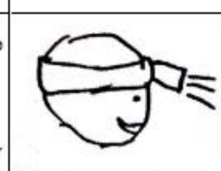
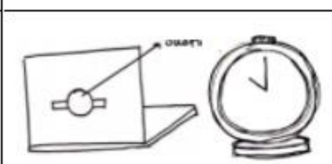
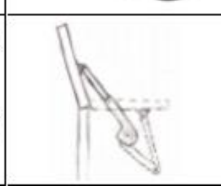
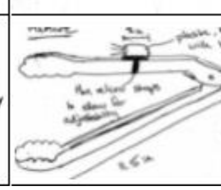
	Line Smoothness		Proportion/Accuracy		Understandability	
Definition	How smoothly (lack of waviness) each line is drawn.		How accurately each line is drawn relative to where it should be drawn.		Can the rater easily understand what the sketcher tried to represent (without relying on words and descriptions)?	
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

Figure 1: 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 independent raters as they scored each page of sketches.

Sketches were assessed by three independent reviewers, including the authors and a graduate design student not working on this project. Raters were given written descriptions of each category and scoring level along with example sketches that would fit each score, as shown in Figure 1, in order to ensure alignment between raters. After an initial round of rating, raters discussed discrepancies and re-rated to ensure that each rater held similar mental models of what the rubric meant. 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."

2.3 Learning from Prototypes

To capture participants' learnings and reflections on the prototypes they created, they were assigned to individually complete an online prototype logging form each time they created a new prototype and were given weekly reminders to continue filling it out for the duration of the project. The form asked for the date the prototype was made, a short text description of the prototype, 1- 5 images of the prototype, and a brief open-ended description of what was learned from the prototype. The goal 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 analysis of the data. The tagging rubric was developed based on existing literature and open-coding through the participant responses. Due to the individual nature of the design project and the remote setting for data collection, the prototypes were primarily "learning" prototypes as defined by Ulrich, et al. [24].

Prior art shows several existing prototype planning tools. Hansen's Prototyping Planner allows designers to plan out prototyping and allow for deliberate decision making [25]. This tool is focused primarily on planning before creating a prototype rather than reflecting on the prototype afterwards. Lauff's Prototyping Canvas scaffolds prototyping planning in much more detail and includes a reflective section on insights gained from testing with the prototype [21]. Since our prototype logger was much more open-ended, we focused on identifying and building on prototype learning outcome categorization schemes that were found in existing literature. Some coding schemas focused on specific prototyping goals such as needfinding [26] or were tailored specifically to the design outcomes of a particular experiment [27]. Others emphasized later stages of the prototyping process than the early-stage design that was the focus of this study [28]. 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 shown in Table 1 and incorporated elements from both Nelson and Lauff's coding schemas and added several unique categories, which will be discussed in the findings [12,13].

Table 1: Each theme and category of prototype assessment along with an example of text that would be tagged under each category. References are included for the categories that are adapted from existing literature.

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	*"Which libraries to install in the Arduino IDE"*	[12]
Planning	Planning next design or fabrication steps	*"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"*	

Three independent reviewers tagged each comment. The final aggregate tags for analysis were determined using the mode of the reviewers' responses (ie. if 2 of 3 reviewers used a tag, it was included and if 2 of 3 reviewers didn't use a tag, it was not included.)

2.4 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 assessed on 7 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 were the mean of the scores given by all reviewers and ranged between 3.3 to 4.7 on the 5-point scale. Scores on this project are subsequently referred to as "design outcomes."

3. RESULTS AND DISCUSSION

3.1 Sketches

Overall sketch quality was measured on a scale of 1-10 as the sum of line smoothness and proportion/accuracy. 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 Figure 1. Metrics such as accuracy of perspective and shading were found in prior art but were not relevant for the vast majority of brainstorming sketches [4,5,9]. An additional category of "understandability" was added to the sketching rubric to assess the ability of the sketch to serve as a communication tool: "Can the rater easily understand what the sketcher tried to represent without relying on words and descriptions?" This was not a category that was found in the prior literature, but is useful to track the effectiveness of the sketch as a communication tool. Additionally, in the virtual setting, participants were more reliant on using sketches to communicate their design ideas to teammates and course staff. Spearman's Rho was used to calculate correlations between attributes of sketches and overall outcomes due to the non-parametric nature of the data.

Krippendorff's alpha for inter-rater reliability was calculated using the open source tool ReCal [29,30]. Sketches were assessed by three independent reviewers with a Krippendorff's alpha of 0.512. Landis and Koch's cut-offs can be used as a benchmarking guideline here to interpret that this alpha value signals moderate agreement between reviewers [31].

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 Figure 2 below. Figure 3 also shows the percentage of students with median sketch ratings at each level. 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 Figure 1. For instance, the bottom right sketch in Figure 1 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.

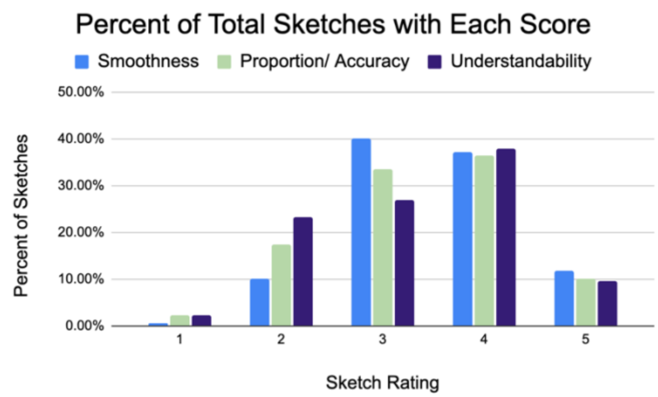


Figure 2: Percent of total sketches rated at each point on a scale of 1-5 (where 5 is the best score) for sketch smoothness, accuracy/proportion, and understandability

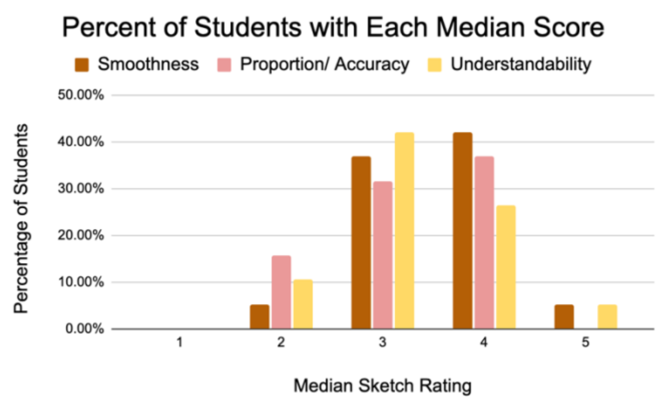


Figure 3: Percent of students with median sketch ratings at each point on a scale of 1-5 (where 5 is the best score) for sketch smoothness, accuracy/proportion, and understandability

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^2=0.224, p = 0.041). This indicates that participants who sketched more frequently had better overall scores on the project. Quantity of sketches also correlated with the perceived innovation of the project (R^2= 0.549, p = 0.015) as rated by reviewers. 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 in order 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 ($R^2 = 0.232$, $p = 0.037$), which may suggest that the more novices sketch, the more likely they are to get at least one "excellent" drawing. Participants with a higher maximum sketch quality score also had higher scores for that design project overall ($R^2 = 0.417$, $p = 0.003$), 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 [4,17]. Surprisingly, these correlations did not persist for the median and means of participant sketch quality, though it is not clear why.

Mean sketch quality and mean understandability were strongly linked ($R^2 = 0.333$, $p = 0.010$), indicating that higher quality sketches were perceived as also being more understandable. Furthermore, mean understandability had strong links with both the mean line smoothness ($R^2 = 0.287$, $p = 0.002$) and mean sketch proportion/accuracy ($R^2 = 0.294$, $p = 0.002$) 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 the other categories (and vice versa). A similar correlation was found for median sketch quality and median understandability ($R^2 = 0.266$, $p = 0.024$) but not for maximum sketch quality and maximum understandability.

3.2 Reflections on Prototypes

As described in the methods section, the prototype logger was used by participants to record each unique prototype made and what was learned from the prototype. 73 prototype logs were submitted by all participants in total. Figure 4 shows an example of images of a project at different stages that were submitted to the logger. One prevalent theme in the reflections on each prototype that wasn't found in existing research literature was the idea of planning for 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 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.

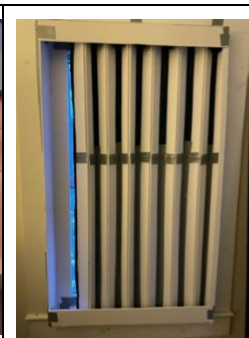
Image		
Description of learned from the prototype	*"The foam core is hard to control when folding together, and doesn't hold its shape when folded out."*	*"While the frame did a good job of acting as a curtain guide, I also expected the front pieces to block light from coming in the top and bottom of 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

Figure 4: 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.

Many subjects included personal reflections or emotional comments as well. These included observations about their own skills, frustrations, and comments that were often unrelated to the actual prototype itself. Finally, many of the 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 inter-rater reliability of 0.633 and average pairwise percent agreement between reviewers of 82.80% [29,30]. This Kappa value and pairwise percent agreement indicate substantial agreement between reviewers [31]. Each overall theme and comment category is listed in Table 1 above along with an example response that would be tagged in that category and 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 2 shows the final 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. Additionally, 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 particular iteration and what was learned from each step in the debugging process. The average entry contained around 36 words.

Table 2 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 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 referenced users even though participants worked with users who were part of their living community so that they could physically test prototypes. It is possible that these novice designers did not leave sufficient time for user testing or felt that their prototypes needed to be finished before testing with users. 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. Similarly, a minimal percentage of comments describe new technical knowledge learned in the process of creating a prototype. As expected, the majority of comments are neutral in tone.

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. Quantity of prototypes logged was positively correlated with percentage of non-neutral (positive or negative) comments (R^2= 0.421, 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.

Table 2: Percentage of comments in the prototype logger corresponding to each category

Theme	Category	Percent of Comments
Build/Test	Engineering performance (assembly or sub-assembly)	86.3%
	Materials	39.7%
	Users	15.1%
Learning	New technical knowledge	5.5%
Planning	Planning next design or fabrication steps	38.4%
Reflection	Personal reflection/ emotion	39.7%
Tone	Positive	17.8%
	Neutral	67.1%
	Negative	15.1%

Though there was a positive trend between the number of logs completed by each participant and overall design outcomes on the project represented by grades in the class and on the project, there was not a statistically significant correlation found. 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, though 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 lack of consistency with logging as a potential source of error in this study.

3.3 Assessment of Designs

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 below in Table 3. Figure 5 shows all the cases in which there is a statistically significant ($p \leq 0.05$) positive correlation between categories.

Table 3: Each category of assessment and its shorthand label for the final design project 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?

Figure 5 shows the R^2 and p-values for the categories that were statistically significantly correlated. Many of the categories had strong correlations with R^2 values of over 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. Alternatively, it is 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). In either case, it is possible that this rubric is over-assessing certain elements of participant designs and should be re-evaluated and modified.

The range (maximum-minimum) of scores in each category was calculated to determine if there were any categories with notable discrepancies amongst 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 variation in the extent to which participants incorporated 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. As seen in the prototyping logger results, very few participants referred to users in the prototyping logs. This points to the importance of emphasizing user feedback to novice designers.

	N	A	TP	I	U	C	PV
N		0.7 *0.0000084*	0.49 *0.00087*				0.61 *0.000072*
A			0.56 *0.00023*	0.34 *0.0083*			0.68 *0.000013*
TP				0.36 *0.0065*		0.34 *0.0085*	0.53 *0.00039*
I						0.24 *0.033*	0.29 *0.016*
U							
C							
PV							

Figure 5: Grid of relationships between design assessment metrics. Cases in which there is a statistically significant positive correlation ($p \leq 0.05$) between categories list the R^2 correlation coefficients (top value) and p-values (bottom value, in italics) in the grid square.

3.4 Limitations

This was a small study with a limited number of participants, and meant to be used as a preliminary first step for other studies. A major limitation of this study as it pertains to prototype logging was that several participants did not log all of the prototypes they created. Additionally, the study was run in the context of an academic course, so it was not possible to have a control group for experiments and we cannot draw causal conclusions.

4. CONCLUSIONS AND FUTURE WORK

RQ 1: *What aspects of early stage design sketches are linked with positive design outcomes for novice designers?*

This study has proposed a framework for assessing quality of "thinking" and "talking" sketches during the ideation phase of the design process. The proposed framework includes categories for line smoothness and proportion/accuracy as well as a new measure of understandability to better assess the effectiveness of sketches as communication tools. We showed that in this case the quantity of sketches correlates with design outcomes (represented here by scores on their design project) for novices, but that 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.

RQ 2: *Are higher quality sketches easier to understand?*
The study examined "understandability" as a criterion for evaluating sketches as it is key for using sketches as a communication tool. In this study, sketch quality is strongly correlated with sketch understandability. This indicates that opportunities for novices to get practice and training in creating sketches may be helpful for them in creating sketches that are understandable and an effective communication tool in the engineering design process. This may have implications for studies around teaming as it is possible that higher quality sketches would be better communication tools between team members and could lead to better designs.

RQ 3: *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 includes 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 which 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 or not intentional reflection on the prototyping process correlates with better outcomes for novice designers.

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. Additionally, it would be interesting to see if similar results hold for practitioners, or if these observations are unique to the novice experience in design.

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 design portfolios are gaining traction in industry and academia. These portfolios often have a focus on showing design process 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 between hand sketches and electronic sketches on tablets as tablets are gaining traction amongst designers. In particular, it would be interesting to note if the correlations between sketch quantity and design outcomes persist when drawings are done on tablets.

Despite 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 both in their prototyping reflections and presentations. This is an area of concern and should be of note to those training novice designers.

5. ACKNOWLEDGEMENTS
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.

REFERENCES
[1] Kudrowitz, B. M., and Wallace, D., 2013, "Assessing the Quality of Ideas from Prolific, Early-Stage Product Ideation," J Eng Design, **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," J Mech Design, **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 Analysis 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."
[6] Rodgers, P. A., Green, G., and McGown, A., 2000, "Using Concept Sketches to Track Design Progress," Design 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," p. V003T04A006-V003T04A006.
[8] Goel, V., 1995, *Sketches of Thought*.

[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 Transactions Interact Intelligent Syst Tiis, **8**(3), p. 19.
[10] Ferguson, E. S., 1994, *Engineering and the Mind's Eye*.
[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."
[12] Nelson, J., Berlin, A., and Menold, J., 2019, "ARCHIE: An Automated Data Collection Method for Physical Prototyping Efforts in Authentic Design Situations."
[13] Lauff, C., Kotys-Schwartz, D., and Rentschler, M. E., 2017, "Perceptions of Prototypes: Pilot Study Comparing Students and Professionals," p. V003T04A011-V003T04A011.
[14] Gerber, E., and Carroll, M., 2012, "The Psychological Experience of Prototyping," Design Stud, **33**(1), pp. 64–84.
[15] Bao, Q., Faas, D., and Yang, M., 2018, "Interplay of Sketching & Prototyping in Early Stage Product Design."
[16] Robertson, B. F., and Radcliffe, D. F., 2009, "Impact of CAD Tools on Creative Problem Solving in Engineering Design," Comput Aided Design, **41**(3), pp. 136--146.
[17] Kwon, J., and Kudrowitz, B., 2019, "The Sketch Quality Bias: Evaluating Descriptions of Product Ideas With and Without Visuals."
[18] 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," p. V005T06A022-V005T06A022.
[19] Häggman, A., Honda, T., and Yang, M. C., 2013, "The Influence of Timing in Exploratory Prototyping and Other Activities in Design Projects," p. V005T06A023-V005T06A023.
[20] Yang, M. C., 2005, "A Study of Prototypes, Design Activity, and Design Outcome," Design Stud, **26**(6), pp. 649–669.
[21] 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.
[22] Menold, J., Jablokow, K., and Simpson, T., 2017, "Prototype for X (PFX): A Holistic Framework for Structuring Prototyping Methods to Support Engineering Design," Design Stud, **50**, pp. 70–112.
[23] 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-V007T06A040.
[24] Ulrich, K. T., Eppinger, S. D., and Yang, M. C., 2020, *Product Design and Development*, McGraw-Hill Education, New York.
[25] 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," Proc Des Soc Des Conf, **1**, pp. 1725–1734.
[26] 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-V007T06A045.
[27] 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."
[28] Lauff, C., Kotys-Schwartz, D., and Rentschler, M. E., 2018, "What Is a Prototype? What Are the Roles of Prototypes in Companies?," J Mech Design, **140**(6).
[29] Freelon, D. G., 2013, "ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service."
[30] Freelon, D. G., 2010, "ReCal: Intercoder Reliability Calculation as a Web Service."
[31] Landis, J. R., and Koch, G. G., 1977, "The Measurement of Observer Agreement for Categorical Data.," Biometrics, **33**(1), pp. 159–74.