

## HOW DESIGN PROPOSALS ARE EVALUATED - A PILOT STUDY

Deiningner, Michael;  
Hansen, Claus Thorp

Technical University of Denmark

### ABSTRACT

In addition to being functional and well-engineered, successful products in today's market are also desirable and appealing to customers. A decision to purchase a product or not is often influenced not only by facts, but also by emotions. Qualities like desirability and appeal that trigger an emotional response can be challenging to satisfy. These qualities can be difficult to quantify and measure and do not easily translate into requirements and specifications. Therefore, understanding the emotional responses of potential customers to product proposals would allow designers to adapt their design strategies. In this study, participants were invited to review a range of design proposals and asked to rank order them relative to one another. The participants were asked to record their comments while discussing their ranking. The findings show that study participants often made decisions about how good or bad they thought a design proposal was without providing rationale to support their rankings. In some cases the rankings were aligned with the comments, but sometimes they were in conflict. More work is needed to further explore the decision-making process and the criteria used when reviewing design proposals.

**Keywords:** Industrial design, Evaluation, Design education

### Contact:

Deiningner, Michael  
Technical University of Denmark  
Mechanical Engineering  
Denmark  
mdein@mek.dtu.dk

**Cite this article:** Deiningner, M., Hansen, C. T. (2021) 'How Design Proposals are Evaluated - A Pilot Study', in *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.433

## 1 INTRODUCTION

Successful products in today's market must offer more than just function. In addition to being well-engineered, they must also be desirable and appealing to customers (Norman, 2004; Seva et al., 2011). Many engineering design programs therefore teach human-centred design competencies extending beyond traditional engineering skills (Böninger, Frenkler and Schmidhuber, 2021). Updated curricula can, for example, include content from humanities and industrial design, and emphasize perception, emotions, colour, textures, etc. This broader content requires ways beyond established engineering principle solutions by which to judge and evaluate design proposals. Ideally, such solutions satisfy a combination of requirements and specifications, including aesthetic features and human sensations (Abidin et al., 2008). However, clearly defining and sharing evaluation criteria for often-subjective design qualities can be challenging. Understanding the emotional responses people have when reviewing design proposals would allow designers to use this knowledge to influence and improve their design proposals, and benefit both students and professional designers working in industry.

This pilot study investigates how design proposals are being evaluated in an attempt to better understand the criteria that effect judgement. What influences our perception when reviewing a design proposal, both in positive and negative ways? What do people focus on when judging new ideas, and can we guide students so that their design proposals elicit positive responses? To advance our understanding of the larger question of "what comprises good design?", this pilot study was conducted to collect and analyse what people focus on when reviewing design proposals. What arguments do reviewers use to support their decisions about whether a design proposal is considered "good" or "bad"?

## 2 BACKGROUND

Customers have many, and increasingly convenient, purchasing options and can choose among a wide variety of products. For new products to be competitive in an ever-expanding global market, they must be well designed, and good design is therefore more important than ever. But what is good design? In engineering design, a focus is placed on satisfying design requirements and specifications. These criteria can be influenced by many factors and, once established, can be evaluated by assessing how well the proposed design solution satisfies them. Many such practices are taught in engineering design programs (Pahl et al., 2007; Dym et al., 2009; Dieter and Schmidt, 2012).

However, the softer requirements like visual appeal or desirability - qualities that can sway a customer to choose one particular product over another - are more challenging to assess. Industrial designers often use tools like mood boards to frame dimensions like aesthetics and appeal of a product. These tools can help to outline a solution space by identifying influential characteristics like competitive products, lifestyle of the target customer, the mood or style of the product to be designed, and product-relevant properties like materials, colours, textures, etc. (Koskinen et al., 2011; Rieuf, 2017).

To help investigate how reviewers evaluate design proposals, we used the Consensual Assessment Technique (CAT), a commonly used tool well-suited to assess and rate creativity or other more subjective criteria of creative works (Amabile, 1983; Amabile et al., 1996; Baer, Kaufman and Gentile, 2004; Howard, Culley and Dekoninck, 2008). With the CAT, subject matter experts are consulted and asked to evaluate creative works (e.g. paintings, essays, etc.) relative to one another and rank order them according to their own criteria. This technique can help to reach consensus and identify reviewers' preferences and dislikes.

## 3 METHODS

In this study, eight product design proposals created by students of Industrial Design 1, a third semester course in the Design & Innovation study line at a technical university were used to collect input from reviewers. Eight groups of reviewers, each consisting of three participants, were presented with collages of 4 images printed on a sheet of A3 paper. These reviewers were provided with the course objectives and learning goals and asked to rank order the eight proposals. The reviewer groups were encouraged to record their arguments and criteria for their ordering in writing. The groups were given 45 minutes to complete this task. Additionally, the grades and comments that an external censor established for each student project during an earlier session were also included in the study and treated as separate data.

### 3.1 Participants

The reviewers in this study included 24 individuals from an engineering design department. They included both female and male reviewers who practiced or taught design and had varying levels of experience judging design proposals. The external censor was an industrial designer with more than 30 years of design and evaluation experience.

### 3.2 Research Design

The design proposals used for this study were projects created by bachelor students who took an optional Industrial Design course as part of an engineering design program. In addition to sketching exercises, the course included lectures on design history, theory and principles. Students who took this course were required to have taken other design courses in earlier semesters, ensuring that all students had some design experience. In the course, the students were free to choose the type of product they wanted to design and were required to hand in a portfolio with the materials they created throughout the semester. Specifically, these materials included a design brief, sketches and physical models to document their ideation process, 3 design proposals, a user scenario, as well as a rendering and physical model of their final design. At the end of the course, the projects were evaluated and graded by an external censor using the Danish educational system grading scale (see Table 1). The design proposals included in this study represent a range of projects that were awarded grades 4 (a fair performance) or higher by the censor.

Table 1. Danish educational system grading scale with ranking point scale

Danish Mark	Explanation	Equivalent ECTS Mark	Ranking Points
12	For an excellent performance	A	52 to 64
10	For a very good performance	B	39 to 51
7	For a good performance	C	26 to 38
4	For a fair performance	D	13 to 25
2	For an adequate performance	E	0 to 12
0	For an inadequate performance	Fx	N/A
-3	For an unacceptable performance	F	N/A

The reviewer groups were presented with A3 boards of the design projects as shown in Figure 1. Each project board included the moodboard, the final sketch of the product, a user scenario, and a photograph of the final physical model built by the students.



Figure 1: Sample A3 project board for a toaster design proposal.

The reviewer groups also received the course and learning objectives for the Industrial Design course and were then asked to order the student projects relative to one another based on how good they thought they were. The reviewer groups were advised to establish their own criteria for evaluating the quality of the design projects.

### 3.3 Data Collection

The written arguments the groups recorded during the ordering activity were collected by the research team, and each group's ranking of the eight project boards (from worst to best design) was captured through photographs. Comments and grades awarded by the external censor were recorded in writing during a separate grading session and included in the study as separate data.

### 3.4 Data Analysis

The reviewer groups were asked to rank order the project boards relative to one another and record their thoughts in writing. This approach was chosen to encourage discussion and negotiation about why one design was ranked higher than another. Through these arguments, it was possible to capture the criteria the groups used to evaluate the design proposals.

First, the rank ordering of the groups was added to an Excel spreadsheet. The highest ranked concepts were assigned 1 point while the lowest ranked concepts were assigned 8 points. The rank numbers of all groups were summarized and a collective score for all 8 projects emerged (see Table 2). Next, the rank orders were translated to grades to allow for a comparison with the grades the censor awarded to the individual projects. To do this, the lowest ranked score was assigned the lowest corresponding grade given by the censor for the selected projects (grade 4), while the highest ranked project was assigned the highest awarded grade (grade 12). The remaining grades for each project were determined accordingly (see Table 1).









Then, the comments the reviewer groups recorded were added to the Excel spreadsheet for qualitative analysis. An inductive coding session (Boyatzis 1998; Creswell 2013; Patton, 2014) was conducted to identify agreements, disagreements and common themes among the groups' comments. The results of

this coding session were used to identify connections among and within the individual groups, among and within the individual design projects and among the groups and the external censor.

## 4 FINDINGS

In this section we present the results of both the qualitative and quantitative analysis. Table 2 shows the individual and combined rankings of the design proposals by the study groups.

Table 2. Rankings of the design proposals by the reviewer groups. From left to right we see the design proposals. The study groups are listed on the left.

	Vacuum	Chain Saw	Compact Camera	Guitar Pedal	Green House	Food Chopper	Folding Camera	Toaster
								
1	6	3	7	5	1	8	2	4
2	4	6	3	8	1	4	1	6
3	3	5	2	7	1	8	6	4
4	5	4	6	8	2	7	1	3
5	4	5	6	8	3	7	2	1
6	8	3	5	1	6	7	2	4
7	5	3	7	8	1	6	4	2
8	5	4	6	8	2	7	1	3
Rank	40	33	42	53	17	54	19	27
Total	5	4	6	7	1	8	2	3

### 4.1 Agreements and disagreements between study groups

The highest ranked project was the greenhouse. Most groups ranked this project as number 1 or 2. One group ranked it as number 3 and only one group ranked it as number 6, indicating strong agreement among the groups. However, despite the high ranking, we recorded few positive comments, and despite their agreement that this was the best design, the groups had more negative things to say about the greenhouse. For example, Group 2, which ranked the greenhouse as number 1, thought the design was "very good," but also added that "the prototype doesn't show the interesting bits." Group 5, which ranked the greenhouse as their number 3 project mentioned that "the quality is high" but also that there is "too much plastic in the model." This group gave the designer credit for "good convincing modelling" but concluded that the design "is not desirable because of the appearance," and that "it is not big enough." Group 7, which ranked the greenhouse number 1, simply commented "good." Perhaps more expected, Group 6, which ranked the greenhouse as its number 6 project, offered only critical comments about the design, including "[it] looks like [an] electric appliance" and criticism that the device was not big enough because "it only feeds one Mojito per 4 weeks of growing."

The lowest ranked proposal was the food chopper. Most groups ranked it as number 8, 7 or 6, and only one group ranked it as number 4. With this lowest ranked proposal, we observed more agreement among the groups than with the highest ranked proposal. Few positive comments were recorded, and Group 7 thought that the "sketches [are] good, [but the] model [is] bad." This group also offered that it saw "some limitations" based on the size of vegetables the device can be used for. Group 2 thought both the "use is questionable" and also commented that the "appeal is very questionable." Group 5 agreed and had this to offer about the appearance: "Weird form and design." Group 5 went even further and thought the design was "not kitchen-like" and instead called it "toilet-like." Outside of visual- and appearance-related comments, several groups were concerned about the safety of the product. Group 3 thought that the "food chopper is too unsafe," and Group 6 summed up the negative comments by positing that the designer had given "no regards to safety," that the device displayed a "weird shape" and that "no need" existed for such a product.









### 4.2 Agreements and disagreements between the study groups and the censor

To enable a comparison of the rank ordering by the reviewer groups with the grading established by the external censor, the rankings were translated to grades as described in 3.2 Research Design (Table 1).



This analysis revealed both agreement and disagreement between the reviewer groups and the censor as shown in Table 3.

Table 3. Grading of product proposals by study groups (Grp) and external censor (Cen)

	Vacuum	Chain Saw	Compact Camera	Guitar Pedal	Green House	Food Chopper	Folding Camera	Toaster
								
Grp	7	10	7	4	12	4	12	10
Cen	7	12	4	4	7	4	12	12

#### 4.2.1 Agreements between groups and censor

The folding camera was awarded the highest grade (12) from the reviewer groups and also received a 12 from the censor. Overall the comments for this design were more positive and several groups praised the quality of the sketches. Group 5 considered this proposal a "handy design" and that it "makes sense to have a reversible monitor." Group 7 echoed this by saying that the "features make it usable / sensible for the lifestyle." But this design proposal also received a few critical comments: Group 2 thought that the design "seems to break if used in the rough" and that the "Mood and styling [were] all over the place." Group 5 agreed with this assessment: "The mood is not matching (indoors is not fitting). Styling: poor match." The censor agreed that the "concepts/final design illustrations are of high quality and explain the concepts very well. Generally, really good sketch work." The censor also thought that the "model shows good details and functionality" and that it "communicates the illustrated design very well."

A design proposal that both the reviewer groups and the censor agreed was not very convincing, and as a result received the lowest grade from both groups, was the guitar pedal. Almost all reviewers voiced concerns about the ergonomics of this product. Group 2 thought the design was "not anatomically sound," Group 3 thought it "would be difficult to use, especially [the] buttons" while Group 6 simply stated "bad ergonomics." Group 5 rated the quality of model and sketches as "poor, not competitive" and Group 2 pointed out that there was "no user [scenario]" and that the design "doesn't match the specification or moodboard." Group 5 seemed disappointed with the novelty of the design proposal: "A round pedal! Nothing new!"

The censor's assessment echoed many of the comments by the reviewer groups, including the ergonomics: "would like more thoughts on user interface - maybe better separation between foot pedal and buttons, this could create conflict/interference as it is now" and "miss seeing an actual foot being part of the exploration during the process - the final design is not as convincing, seems too small." The censor also thought that the "imageboard puts you in the right mood but doesn't really reflect the final design" and "a fair design process even if the sketches evolve a lot around the same forms." Finally, the censor criticized the level of detail: "model and illustrations are fine but the detailing could be explored better."

#### 4.2.2 Disagreements between groups and censor

We also identified a design proposal where the reviewer groups' appraisal deviated significantly from that of the censor: the greenhouse. The reviewer groups gave this project a 12 while the censor only gave a 7, a 2-grade difference. While the reviewers had few positive things to say about this proposal to support their high grade, the censor's assessment was more balanced with both positive and negative comments: "Nice executions of the model and good illustrations," but "a little doubtful about the chosen design solution - does it contribute anything to the plant growth?" The censor further commented that "a combination of the features from concept 1 regarding artificial light, and water storage/grove from concept 2 would have made the final design more convincing - a shame they weren't chosen for the final design." This was summarized by "Good process with many good ideas/thoughts - just a shame a lot of the initial ideas didn't make it to the final design."

#### 4.3 Conflicting statements & disagreements within reviewer groups

Several groups were concerned with the safety of some of the design proposals, most noticeably the chain saw and the food chopper. For example, Group 1 questioned the "safety" of the chain saw;

Group 2 thought that the round shape would lead to a more "unsafe" product; Group 3 worried about "conformity with regulations" and also called the design "unsafe." Group 7 stated that the design was "lacking safety but was otherwise good." Group 3 thought that the food chopper was "too unsafe" and Group 6 felt that the designer had given "no regards to safety" when designing the product.

When judging the appearance of the two products, not all groups agreed and offered comments like "not elegant or attractive" (Group 5) but also "the round shape is attractive" (Group 6) for the chainsaw. Similarly, we recorded controversial comments for the food chopper. Group 2 thought that "the appeal is very questionable"; Group 5 offered "weird form and design" while Group 7 thought that the "shape and finish are ok." Although groups shared cohesive concerns regarding safety for both products, they were divided on how they felt about the visual appeal. Overall the chainsaw was ranked as number 4 while the food chopper was ranked lowest as number 8.

The lowest ranked item by Group 6 was the vacuum cleaner. The group thought that it was "simple" and offered strong words like "we hate it" and "what's up with the ergonomics?", comments that justify its low positioning of this product. Overall the vacuum cleaner was ranked number 5 by all groups.

Group 6 gave its highest score to the guitar pedal, the item that scored second lowest on the overall ranking list. The group provided positive arguments such as "[The] Retro [design theme] is translated with promise" but also argued that the design is "incomplete" and possessed "bad ergonomics," making it challenging to understand why it ranked this product as the best design.

Additionally, Group 6 ranked the greenhouse, the product that received the overall highest score by all groups, as number 6 and only offered negative comments such as "looks like an electrical appliance" and "only feeds one Mojito per four weeks of growing." Here, the negative comments reflected the low score but were in conflict with how the majority ranked the same product. For example, Group 2 called the design "very good" and Group 5 thought that "the quality is high."

## 5 DISCUSSION

### 5.1 Agreements and disagreements between study groups

We observed groups for which their ranking was supported by their comments. For example, when Group 5 commented on the food chopper, which it ranked second lowest, the group offered mostly negative comments that matched its low ranking: "Weird form and design, not kitchen-like, toilet-like. The application, the way you hold it is not desirable. The mood is nice but not relevant to the design." Similarly, this group offered rather positive comments for the toaster, which they awarded their highest ranking: "Models are nice, a little crowded. It is recognizable - interesting and attractive. The size of bread is problematic. The mood is good. Colours match." Group 6 also had comments consistent with their rankings: for the vacuum cleaner, their lowest ranked design, the group offered "Simple, we hate it! And what is up with the ergonomics?"

Some of the comments that the groups provided did not match well with their rankings, making it more challenging to understand their rationale. Group 6 for example offered controversial comments for the guitar pedal, their highest ranked project: "Retro is translated with promise, but incomplete. Bad ergonomics." Similarly, the greenhouse was Group 7's favourite design but only received a "good" for the sketches and "we see the link" with the moodboard. This group thought that the guitar pedal was the least desirable design but offered comments such as "Good but missing perspectives / views and the use" and thought that the moodboard was "not bad," comments that do not directly support the group's low ranking.

Similarly, Group 2 ranked the greenhouse the highest, but then had only negative comments to offer, like: "The prototype doesn't show the interesting bits." In order to still rank the design high, the team might have assumed, and trusted, that the designer worked out the details required for it to be a well-functioning item, when in fact it was impossible to evaluate this through either the model or the sketches.

All the groups' comments were better aligned on the design proposals that received low rankings than on the proposals with high rankings. This might suggest that our reviewer groups were better prepared to offer critique rather than to give praise when judging design proposals (Peale, 2012).

## 5.2 Disagreements between the study groups and the censor

We observed the most noticeable disagreements between the groups and the censor for the greenhouse. The reviewer groups had only a few positive things to say about this design proposal and provided mostly negative comments. However, without justifying their decision or providing an explanation, they ranked this design highest (12). In contrast, the censor was more critical, provided both positive and negative comments, and assigned a grade that better reflected this assessment (7).

The censor had the benefit of reviewing more of the design process in the form of a portfolio that included all the material the students created throughout the course. This provided access and insight into other ideas and thoughts that did not translate into the final design, and this additional information could have influenced the assessment. Due to the time constraints in this study, the reviewer groups were only shown a selection of this portfolio, however, the selection included the most significant pieces, namely the final deliverables.

It is unclear if less information led to the often unsupported yet positive evaluations. It is possible that the reviewers assumed that essential details such as how the greenhouse supports plant health were indeed resolved by the designer even when this was not evident. However, this information was also not transparent to the censor since this particular student did not explore this detail anywhere. It is therefore unclear what led to some of the positive rankings by the reviewers as they did not offer comments that supported their decisions, nor did they state a lack of information, and both positive and negative rankings were unjustified.

## 5.3 Conflicting statements & disagreements within study groups

Some of the proposed products prompted the reviewers to comment on safety. These comments were reserved for design proposals with an apparent risk of injury, like the chainsaw or the food chopper—products that either cut or chop items and could harm a user if safety precautions were not considered in the design. Even though there are potential safety concerns with other products as well (for example burning or electrocution with the toaster), reviewer groups did not comment on those. And while it may be obvious that one could get hurt when using a chainsaw, it is not clear if the proposed design would be any better or worse in regards to safety when compared with products that are on the market today. Nevertheless, half of the groups voiced safety concerns for the chainsaw. This could be due to the fact that the safety risk of using a chainsaw is generally high, and not necessarily only for this particular design proposal. On the other hand, the model and illustrations of the food chopper did not explain if the designer considered safety. It is unclear if there are any features that would prevent a user from getting injured, yet only two groups commented on safety for this design proposal. The perceived risk of injury with a food chopper might be lower than that of a chainsaw. It is therefore possible that the perceived safety risks of a product category could have influenced the reviewers' opinions more than the details of a specific design proposal. However, in spite of the safety concerns many groups had about the chainsaw, this design proposal was ranked much higher (number 4) than the food chopper, the lowest ranked design proposal by all groups. The chainsaw was regarded as more appealing than the food chopper, and perhaps the visual appearance was more important to the reviewer groups than their safety concerns. This aligns with research that found that reviewers rate products they consider more beautiful as being also more useable (Tractinsky, Katz and Ikar, 2000; Tuch et al., 2012). Whatever the reasons for their rankings, they were not transparent from the reviewers' comments.

## 6 LIMITATIONS

This study included the limited number of reviewers with similar backgrounds and we restricted the time these reviewers had to evaluate the design proposals to 45 minutes. Further, the reviewers were only presented with selected project deliverables while the external censor had access to all the material the students created. Translating the rank order of the reviewer groups to a grade scale for comparison with the evaluations of the external censor might not have resulted in the same grades the reviewer groups would have assigned themselves had they been asked to do so instead of rank ordering the projects.



## 7 CONCLUSIONS

The results of this study show that reviewers often made decisions about how good or bad they thought a design proposal was without providing rationale to support their rankings. Some rankings aligned with the comments, some were in conflict. For example, positive or negative comments on safety did not always lead to a corresponding ranking. Therefore, other criteria that were not articulated by the reviewer groups must have influenced the decision-making that led to the final rankings. Perhaps only the most significant and unique comments were recorded - "the squeaky wheel gets the oil" - while the groups based their decisions for ranking of a design proposal on other criteria that were not recorded. More work is needed to further explore decision-making and the criteria used when reviewing design proposals.

## ACKNOWLEDGEMENTS

The authors thank the study participants for their time and effort to review and rank the design proposals, the students who created these design proposals, and the conference reviewers for their input.

## REFERENCES

- Abidin, S. Z. et al. (2008) 'On the role of formgiving in design'. <https://dx.doi.org/10.13140/2.1.1922.4649>.
- Amabile, T. M. (1983) 'The social psychology of creativity: A componential conceptualization', *Journal of Personality and Social Psychology*, 45(2), pp. 357–376. <https://dx.doi.org/10.1037/0022-3514.45.2.357>.
- Amabile, T. M. et al. (1996) 'Assessing the work environment for creativity'. *The Academy of Management Journal*.
- Baer, J., Kaufman, J. C. and Gentile, C. A. (2004) 'Extension of the Consensual Assessment Technique to Nonparallel Creative Products', *Creativity Research Journal*, 16(1), pp. 113–117. [https://dx.doi.org/10.1207/s15326934crj1601\\_11](https://dx.doi.org/10.1207/s15326934crj1601_11).
- Böninger, C., Frenkler, F. and Schmidhuber, S. (2021) *Designing Design Education – Whitebook on the Future of Design Education*. iF Design Foundation.
- Dieter, G. and Schmidt, L. (2012) *Engineering Design*. 5 edition. New York: McGraw-Hill
- Dym, C. et al. (2009) 'Engineering Design: A Project-Based Introduction', All HMC Faculty Books. Available at: [http://scholarship.claremont.edu/hmc\\_facbooks/24](http://scholarship.claremont.edu/hmc_facbooks/24).
- Howard, T. J., Culley, S. J. and Dekoninck, E. (2008) 'Describing the creative design process by the integration of engineering design and cognitive psychology literature', *Design Studies*, 29(2), pp. 160–180. <https://dx.doi.org/10.1016/j.destud.2008.01.001>.
- Koskinen, I. et al. (2011) *Design Research Through Practice: From the Lab, Field, and Showroom*. Elsevier.
- Norman, D. (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*. (New York: Basic Books.
- Pahl, G. et al. (2007) *Engineering Design*. 3rd edn. Springer London.
- Peale, N. V. (2012) *The Power of Positive Thinking*. Random House.
- Rieuf, V. (2017) 'Emotional activity in early immersive design: Sketches and moodboards in virtual reality', 48, p. 33.
- Seva, R. R. et al. (2011) 'Product design enhancement using apparent usability and affective quality', *Applied Ergonomics*, 42(3), pp. 511–517. <https://dx.doi.org/10.1016/j.apergo.2010.09.009>.
- Tractinsky, N., Katz, A. S. and Ikar, D. (2000) 'What is beautiful is usable', *Interacting with Computers*, 13(2), pp. 127–145. [https://dx.doi.org/10.1016/S0953-5438\(00\)00031-X](https://dx.doi.org/10.1016/S0953-5438(00)00031-X).
- Tuch, A. N. et al. (2012) 'Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI', *Computers in Human Behavior*, 28(5), pp. 1596–1607. <https://dx.doi.org/10.1016/j.chb.2012.03.024>.



CAMBRIDGE  
UNIVERSITY PRESS