

CAN COMPARATIVE JUDGEMENT IMPROVE PRODUCT DEVELOPMENT AND PRODUCT-SERVICE SYSTEM DESIGN FOR STUDENTS

Ivo DEWIT¹, David CORRADI² and Maarten GOOSSENS³

¹University of Antwerp, Faculty of Design Sciences, Department of Product Development

²University of Antwerp, Faculty of Design Sciences, Dean's Office

³University of Antwerp, Faculty of Social Sciences, Department of Educational Sciences

ABSTRACT

One issue within current product development literature entails, both with students and with professionals, the reliable scaling, and the objective assessment of products. This article will attempt to describe a way how master's students at the University of Antwerp, within a product development case, can both assess each other, assess the process of product development, and assess the final product using comparative judgement (CJ). This, while achieving learning goals described at master level. Our evaluation was anchored by use of Comproved, a validated instrument designed to comparatively evaluate products of students, during the design process (formative) and when uploading the final deliverables (summative) for the course. The current study evaluates how first year Master students (n=72) apply CJ using an online tool, Comproved (Comproved.com) during a semester-long group project on Product-Service System (PSS) design. Based on post hoc questionnaire details and interviews, we will list the advantages and the disadvantages of CJ, as stated by students. Finally, we will list the potentials of CJ during product development cycles and their impact on product quality, reliability of judgement and metacognitive strategies for learning.

Keywords: Design education, design process, product-service systems, reliability, assessment, comparative judgement, metacognition, and learning

1 INTRODUCTION

Product-Service Systems (PSS) are systems in which products and services are designed in unison, reinforcing each other, putting emphasis on the whole and how they work together rather than being designed separately. But putting emphasis on its integral nature, PSS design requires a structured process with a broader scope in mind. Tools and methods to support PSS design exist in different formats. However, there is a big difference in having the right tools, having the competences to act on them and being able to question the overall approach independently [1, 2].

In a learning context, students are taught the 'PSS Design Toolkit' [3] which provides an opportunity to create innovative interactions between consumers, the products and services they use, and the providers offering them. In cognitive terms, learners require to make links and translate between different levels and aspects of product development, to account for all interacting product aspects. This, while at the same time never losing sight of the 'big picture'. This requires advanced analytical thinking (understanding every small part of a product or service), and synthetic competencies (how the parts lead to an innovative design), and metacognitive awareness (why a combination of subparts leads to a better service, while another combination doesn't), e.g., Puryear (2015).

In contrast to traditional product design, the main challenge for student designers is coping with this increased complexity, i.e., a multitude of interacting variables. Additionally, they are required to manage a variety of underlying design processes, where early representation and communication are key for the design phase and enable a more successful strategic rollout of the PSS concept. With our toolkit, we posit that reiterative exploration is critical to support the designer's creative process, but are those processes alone helping students see 'the bigger picture'?

2 COMPARATIVE JUDGEMENT

A growing body of literature is finding support for the notion that comparative judgment can help learners and assessors in different learning and working situations [5, 6]. Comparative judgement (CJ) asks an assessor to compare two products and ranks one product over the other. These products can be both small scale – such as a short presentation, a paper, or a drawing; and large-scale such as a masters’ thesis or a full fledge solution to a real-life solution to a problem. Doing multiple comparisons using multiple judges (students, teachers or external assessors), a measurement scale can be created using the Bradley-Terry-Luce model [7, 8] showing the relative quality of each product [9]. Furthermore, each product receives an ability score and detailed feedback.

A major strength of CJ, in terms of assessing difficult-to-specify constructs such as a product design process, is that the result is based on the collective expertise of the rater. Or, its validity is anchored in what is valued by the community of practice within a given discipline [10].

Studies indicate CJ can be beneficial in some cases during a learning process [11, 12]. Where rubrics, even those which specify performance levels and/or standards for each criteria, are often far too abstract for students to really grasp what ‘quality’ is for a given competence [13], CJ shows exemplars to students which is more relevant when it comes to understanding what quality is [14–16]. By analysing exemplars, students experience how high-quality products differ from average products [17]. Furthermore, by comparing products, learners are taught to think why product a is better than product b and learn to articulate why a product is better. This thinking process is often referred to as metacognition, whereby the learner attempts to understand all intricate aspects why a product is better than another one. When learners compare each other’s products or services, both during and at the end of an educational process, learners are also comparing each other’s analytical and synthetical thinking process, which can improve their metacognitive awareness of the product development process.

Compared to using rubrics, the comparative assessment is done in terms of holistic processes (Figure 1), whereby the whole is more than its aspects, or when it comes to designing with increased complexity, e.g., product-service systems (PSS), resulting in a more reliable grading for complex assignments [6, 18].

Use both lenses to assess the video (also mention the perspective)

User perspective (are you at the center of attention? are the insights relevant? is this context yours? is this your problem? ...)

Type your feedback for product A here ...

Client perspective (is this what you paid for? have they made progress? were they able to summarize the insights | content-wise & graphically? ...)

Type your feedback for product A here ...

Feedback criteria

Submit

User perspective (are you at the center of attention? are the insights relevant? is this context yours? is this your problem? ...)

Type your feedback for product B here ...

Client perspective (is this what you paid for? have they made progress? were they able to summarize the insights | content-wise & graphically? ...)

Type your feedback for product B here ...

Figure 1. An example of a comparative judgment task where students choose either A or B as a better product and provide feedback based on certain criteria

Not all studies found all-round positive effects. Bartholomew et al. (2019) for example found that feedback given by students in a CJ setting was rather superficial and mainly aimed at the aesthetics of a certain design, and rarely in-depth. Other studies have indicated that not just any product can be compared. Comparisons with too much information, or which are too different, e.g., [19], do not lead to reliable results.

2.1 Research Question

Rubrics usually divide the product into small aspects, a presentation is subdivided into criteria on layout, on persuasive message, and so on, where each criterium receives a separate score. CJ, on the other hand, focusses on the entire product and compares the entire product as such into a holistic process [6, 18]. In the following paragraphs, we will detail the relevance of this holistic assessment in the context of product development education.

While previous research has found both a myriad of positive and negative educational effects using CJ, our study aims at understanding how students perceive CJ during such a PSS design process. Do students feel CJ is a valuable tool? Do students feel it has improved their metacognitive thinking? And finally, what are perceived strengths and weaknesses by students when using CJ during a learning process and for the final assessment.

3 RESEARCH METHOD

3.1 Educational setting

Seventy-two master students in product development were placed into groups of four. Most students did not have experience with the PSS design toolkit in the design course and do not have experience with designing PSS. The program curriculum and this particular course was until very recently taught in Flemish, which limits our student population to those from Belgium, the Netherlands, and still a limited number of European students (Erasmus mobility). Gender is approx. equally divided.

This population participates in the design course entitled 'IPO Project: Integrated Systems'. To achieve the objectives for this twelve-week design course, student teams are expected to generate user insights, explore new opportunities, and then define and design a relevant Product-Service System (PSS) concept for a given context, such as 'milestones in life', 'physical and digital traces', 'tackling obesity', 'smart city sharing', 'Quadruple Helix' and specifically this year's topic 'Future Urban Health'.

During the course we used a CJ tool, Comproved, to allow students to give feedback and rank each other's weekly presentation, mid-term, and final presentations. This presentation was a short, personalized visualization of their project (ranging from two to ten minutes), which allowed for a comparison between the different student teams, their design process and progress. In the middle and at the end of the design course, a jury was held evaluating the projects. The evaluative role of CJ in here was both formative (weekly and mid-term presentations) and summative (for the final presentation).

3.2 Participants and Instruments

We used a mixed method methodology to broadly understand how CJ has impacted the students' learning process. We asked participating students to provide a reflection of half a page (max.) on the general educational experience. About all of the students (n=72) sent in the information. An additional survey was sent, asking broad and specific questions on educational quality (n=18) [20], to have a more detailed insight into the educational experiences using CJ. To avoid bias, we did not ask specifically about Comproved in either the reflective questions or the survey.

4 RESULTS AND DISCUSSION

4.1 Qualitative analysis

Two raters independently categorized results of the open questions into 3 broad categories: strengths, mixed/neutral experiences, and weaknesses of CJ. Afterward, ratings that did not have a consensus were discussed and recategorized. We found 12 remarks that related to the strengths, 9 remarks that were categorized as neutral/mixed and 4 remarks that were perceived as weaknesses. In this paper we will focus on the positive (strengths) and negative (weaknesses) educational effects as indicated by the students using CJ.

4.1.1 Strengths of comparative judgement

Students understand that serious evaluation of these products is important. The CJ software was nice to work with and to receive feedback. Especially the videos via the software worked very well, students often got good feedback from them. They experienced it as interesting to give feedback like this and stated: *"If we had given a live presentation in an auditorium for the mid-term, we would not have received as much feedback as now"*. Overall, the students are interested to follow the processes of the other groups, making it fun and informative to see what others are working on and how they approach things. CJ indirectly supports the process and delivers overall better presentation quality of the students' products: *"I found it very instructive to make a video to show our progress as a group, because we learned to present better (a) but this was also the ideal moment to critically review our work (b) while making the video"*. *"Making the short interim progress videos helped to motivate us (a) and to prepare us for creating the final video(s) (b), and the other videos were also inspiring for our own work (c)"*.

“The videos seemed to me to be more work than a regular morning presentation of the project, but like this we learned to work better with *première pro*, resulting in a new skill for the portfolio”. The CJ software provides a dashboard with all the other products so you can explore them on a later occasion. Therefore, CJ proved useful for groups who want to see the progress of other groups (a), and to get their own progress and presentation skills assessed (b). Furthermore, the CJ software offers the ability to open up the rankings so that both place (quality rating) and feedback on other work is available and that this is perceived as a great learning opportunity (Figure 2).

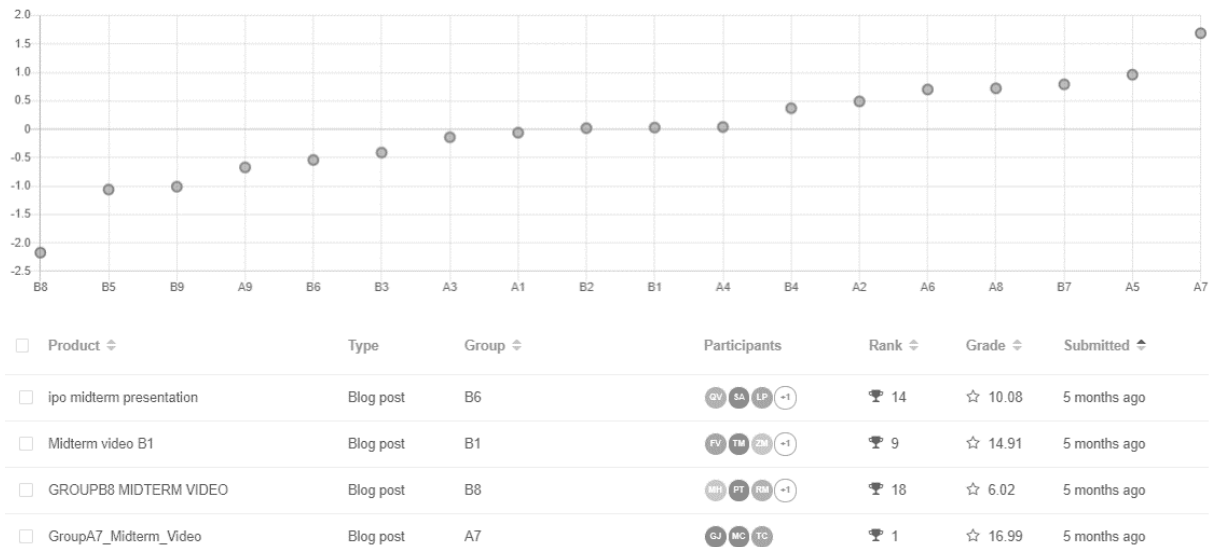


Figure 2. The comparative judgment software allows students to view and compare their rank (grade when selected by the teachers) and feedback with other work

4.1.2 Weaknesses of comparative judgement

For interim products, student preferences go to the small-scale short presentation, a paper, or a drawing. When it comes to the large-scale products – describing full fledged solutions to a real-life problem (PSS design) – students prefer to compare and assess fewer products, people would watch it more carefully, focus better and give better feedback. From a certain length, it felt like only a part of the video (product) was viewed each time, resulting in negative reactions from people who did not understand, or gave comments that were totally irrelevant, not useful, or at least much less than the students (receiving the feedback) expected. Consequently, we have to be more precise on communicating the assessment criteria and goal of the assessment itself: “Rating is so black and white, sometimes the idea of product a was better than product b, but the product (video) was worse. It felt like comparing apples with pears”. Students also perceived a certain difficulty with the overview of comments, there is no hierarchy or difference when stating the feedback on the product being from professors or students. This resulted in not knowing what was good and what certainly was not good (enough) in terms of prioritization and being able to proceed. This forced us to talk the feedback over in the consecutive group consult, making space for a discussion between student and teacher, in order to summarize CJ feedback and emphasizing priorities opposed to confusing or irrelevant comments. Finally, we have to take into account, that CJ products should always be uploaded earlier (putting more pressure on the deadline), and making a video takes much more time than e.g., preparing a presentation.

5 CONCLUSIONS

This article describes how master’s students of Product Development can assess each other, the progress during the design, and the final ‘product’ using CJ. The results show that the students perceived CJ to be helpful during and at the end of the design process. Furthermore, we describe a crossover between CJ and PSS design learning objectives:

- CJ motivates students to envision early stage conceptualizations and representations of the design (process), as such new prospects are opened up, envisioning possibilities and next steps.
- A continuous comparison of interim ‘products’ supports student designers to explicit their doing, making, and inventing, enabling communication with others and provide a better understanding of

- the design process, which otherwise would remain largely tacit knowledge (even for the designer).
- CJ allows streamlining of the design process, makes it consistently comprehensible and provides a benchmark between peers. Specifically for PSS design, designers should be apt to deal with communicating its complexity and achieve simplification with the right representation.
- Unless provided, a set of alternative solutions - to compare with - is usually not a given. CJ allows comparisons between designs in terms of better and worse, but also in judgement of best.

This article has tested the CJ tool in the context of a complex (i.e., real-life) PSS design assignment. Besides the measurement concerning students' face validity with the instrument, the presented results of the CJ software show stability and reliability in two ways:

1. *Power in numbers*: 6 times 24 users (weekly basis) (n=144) and in the middle of that 1 time 72 users for the formative use of CJ, and in the end, again data from 72 users for the summative use of CJ provides significant reassurance that the results we have presented are meaningful.
2. *Reliable scaling*: To compare the same products by multiple raters results in a more objective assessment of the products compared to more subjective grading using rubrics (criteria and interval scoring) of design products by their professors. This, because comparing is a more natural way of assessing [21] in which people tend to be better than in making absolute judgements [22].

We have also corroborated the CJ results together with a rubrics evaluation in order to create a more detailed and nuanced picture of the results. As such, CJ results help frame a more cohesive story for the support in peer-assessor evaluation and student engagement in that process.

When initiating this course, the design process, its progress, and intermediate presentations would normally take place on campus. However – due to COVID-19 – the educational program passed rather quickly into in 'code red', forcing us to completely rethink our remote approach. At that time, Comproved was being integrated in the digital learning platform of the University, ready to be pioneered. Rather by forced coincidence, this CJ software showed great strength as an online substitute to the formative and summative evaluation. It proved to be a direct addition and support to the design process and its progression, something that we will definitely hold on to in post-COVID teaching times.

6 FUTURE RESEARCH

While CJ has proven to adding to the understanding of evaluation as part of students' active education in this one course, we would still like to see how CJ supports the creative work and critical reflection of students graduating and becoming design professionals. Furthermore, there is a difference between student usage of CJ and its effects on the assessor evaluation of a design course in an educational setting. Therefore, our avenues for future research will gear toward usage of CJ to assist in a more objective and reliable evaluation, and see this research extended by design academics evaluating their more complex design courses at other educational institutions (using Comproved or similar CJ software).

REFERENCES

- [1] Dewit I. (2019) *Product-service system design, a synthesis approach*. University of Antwerp
- [2] Dewit I., Jacoby A. and Matthyssens P. (2021) *Design Preconditions for Product—Service Integration*. *Designs* 5:. <https://doi.org/10.3390/designs5020029>
- [3] Dewit I., Van Ael K. and De Roeck D., et al. (2018) *PSS Design and Strategic Rollout: tools for product-service systems*. University Press Antwerp (UPA), Antwerp
- [4] Puryear J. S. (2015) *Metacognition as a Moderator of Creative Ideation and Creative Production*. *Creat Res J*. <https://doi.org/10.1080/10400419.2015.1087270>
- [5] Lesterhuis M., Verhavert S., Coertjens L., et al (2016) *Comparative judgement as a promising alternative to score competences*. In: *Innovative Practices for Higher Education Assessment and Measurement*
- [6] Van Gasse R., Mortier A., Goossens M., et al (2017) *Feedback opportunities of comparative judgement: An overview of possible features and acceptance at different user levels*. In: *Communications in Computer and Information Science*
- [7] Bradley R. A. and Terry M. E. (1952) *Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons*. *Biometrika* 39:. <https://doi.org/10.2307/2334029>
- [8] Luce R. D. (2005) *Individual Choice Behaviour: A Theoretical Analysis*, 2nd ed. Dover Publications
- [9] Pollitt A. (2012) *The method of Adaptive Comparative Judgement*. *Assess Educ Princ Policy Pract* 19:. <https://doi.org/10.1080/0969594X.2012.665354>

- [10] Jones I., Swan M. and Pollitt A. (2015) *Assessing Mathematical Problem Solving Using Comparative Judgement*. Int J Sci Math Educ 13:151–177. <https://doi.org/10.1007/s10763-013-9497-6>
- [11] Bartholomew S. R., Zhang L., Garcia Bravo E. and Strimel G. J. (2019) *A Tool for Formative Assessment and Learning in a Graphics Design Course: Adaptive Comparative Judgement*. Des J. <https://doi.org/10.1080/14606925.2018.1560876>
- [12] Bouwer R., Lesterhuis M., Bonne P. and De Maeyer S. (2018) *Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing*. Front Educ 3:. <https://doi.org/10.3389/educ.2018.00086>
- [13] Brookhart S. M. (2018) *Appropriate Criteria: Key to Effective Rubrics*. Front. Educ. 3
- [14] Boud D. (2000) *Sustainable assessment for long term learning*. Stud Contin Educ 22:
- [15] Nicol D. and MacFarlane-Dick D. (2006) *Formative assessment and self-regulated learning: A model and seven principles of good feedback practice*. Stud High Educ 31:. <https://doi.org/10.1080/03075070600572090>
- [16] Carless D. and Boud D. (2018) *The development of student feedback literacy: enabling uptake of feedback*. Assess Eval High Educ 43:. <https://doi.org/10.1080/02602938.2018.1463354>
- [17] Orsmond P., Merry S. and Reiling K. (2002) *The use of exemplars and formative feedback when using student derived marking Criteria in peer and Self-assessment*. Assess Eval High Educ 27:. <https://doi.org/10.1080/0260293022000001337>
- [18] Goossens M. and De Maeyer S. (2018) *How to obtain efficient high reliabilities in assessing texts: Rubrics vs comparative judgement*. In: Communications in Computer and Information Science
- [19] Slovic P. and MacPhillamy D. (1974) *Dimensional commensurability and cue utilization in comparative judgment*. Organ Behav Hum Perform 11:. [https://doi.org/10.1016/0030-5073\(74\)90013-0](https://doi.org/10.1016/0030-5073(74)90013-0)
- [20] Spooren P., Mortelmans D. and Denekens J. (2007) *Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales*. Assess Eval High Educ 32:. <https://doi.org/10.1080/02602930601117191>
- [21] Laming D. (2004) *Human Judgment: The Eye of the Beholder*. Thomson Learning
- [22] Thurstone L. L. (1927) *A law of comparative judgment*. Psychol Rev 34:. <https://doi.org/10.1037/h0070288>