Design and Technology Education: An International Journal

26.3

# How can comparative judgement become an effective means toward providing clear formative feedback to students to improve their learning process during their product-service-system design project?

**Ivo Dewit, University of Antwerp, Belgium**
**Sarah Rohaert, University of Antwerp, Belgium**
**David Corradi, University of Antwerp, Belgium**

## Abstract

This study responds to calls to further investigate ways to make feedback more effective for students in the context of higher education. More specifically it scrutinizes the feedback practice, adapted to the exceptional reality of a partly on Campus, partly online semester–long Product-Service System (PSS) design project for first Master students of X at the University of Y. To do so, an established model of feedback (Hattie & Timperley, 2007) is used as a framework to seek answers to the research question: which types (and levels) of feedback are generated when applying Comparative Judgement (CJ) to guide the students' and teachers' feedback formulation? Following the model, first three types of feedback: feeding up; feeding back; and feeding forward and second, four levels of feedback are discerned: task; process; self-regulatory and self. The current study describes how first year Master students (n=72) and lecturers (n=4) apply CJ to formulate feedback. We evaluate which types and levels of feedback are formulated and received by the students, both towards and from their peers and teachers. Additionally, based on a post hoc survey and reflection paper, we list the strengths and weaknesses of CJ as a method to help students to formulate, interpret and receive feedback. Finally, we identify various opportunities to improve CJ based feedback during product development cycles and its impact on learning and self–assessment of the own project process and (intermediate) results quality, and metacognitive strategies for learning.

## Keywords

design education, formative feedback, product-service systems, comparative judgement, metacognition and learning

## Introduction

Kluger and De Nisi's (1996) meta-analysis of feedback was a catalyst for acknowledging and investigating the highly variable effects of feedback on learning, both positive and negative. Further research demonstrated that a large amount of feedback by teachers is rarely used and implemented by students (Carless, 2006). This study responds to calls to further investigate ways to make feedback more effective for students (Brooks et al., 2019; Hattie et al., 2016; Shute, 2008). Hereto we explore if and how Comparative Judgement (CJ), i.e., comparing and ranking 2 or more products, can support the formulation of effective formative feedback. We do so in the context of a partly on campus, partly online semester–long Product-Service System

(PSS) design project for first master students of X at the University of Y during the exceptional fall of 2020, profoundly marked by the second wave of the Covid-19 pandemic.

## Teaching and learning context: Product-service systems

This paper discusses the findings of a research study focusing on the question of whether CJ can support the formulation of clear, actionable formative feedback to guide the students' learning process. Specifically, this question will be framed within the context of guiding a PSS design learning process. Now what does a PSS design project entail and which kind and frequency of feedback does it call for? PSS design projects seek to create systems in which products and services are designed in unison and reinforce each other. As such, PSS design requires a structured process with a broad scope to embrace the integral nature of complex system–level challenges and potential solutions. As such, it fits well with a design pedagogy vision, supported by Tovey (2015), who claims that designers should become generalists in as wide a range of content as possible, as the wider the reach of their knowledge base, the more likely the creative inspiration to address complex challenges. This will allow designers to work on multifactorial and global challenges, as well as it will help them to address dynamic problems that evolve as design projects develop (Rittel & Webber, 1973). This view is supported by theories about critical reflection in critical dialectic approaches (Habermas, 1978), which requires students to incorporate multiple perspectives into their thinking and eventually into their designs.

To incorporate multiple perspectives throughout their design process, the students are taught the 'PSS Design Toolkit' (Dewit et al., 2018) which provides the tools to observe, interact with and receive feedback from various stakeholders and guidance on which tools to use towards which ends in which design phases. This toolkit supports the students to gradually create innovative interactions between consumers, the products and services they use, and the providers offering them. However, there is a big difference in having access to the right tools, having the competences to use them, and being able to question the overall approach independently (Dewit, 2019; Dewit et al., 2021). In cognitive terms, learners are required to make links and translate between different levels and aspects of the integrated product and service development process, while considering all interacting PSS aspects and actors. At the same time the learners do have to take care not to lose sight of the 'big picture', the integrated whole.

From the part of the learners, advanced analytical thinking (to understand every small part of a product or service), and synthetic competencies (to understand how all parts combined can lead to an innovative design) are required, just like metacognitive awareness (Puryear, 2015) (to understand why a certain combination of subparts leads to a better service, while another combination doesn't) (Medola et al., 2021) Receiving adequate, clear and regular feedback, both from project stakeholders and from teachers and peers is key to further develop above mentioned competencies (Callender et al., 2016), as these are necessary to act on new insights and reframe their thinking.

Therefore, from the part of the teachers, the scheduling of frequent checkpoints throughout the learning period gives students multiple opportunities to demonstrate their learning and project progress. These checkpoints provide teachers with an information of how their students are proceeding towards achieving the learning goals (Brooks et al., 2019). Furthermore, they

provide various opportunities to give formative feedback to guide the learning process. The most important activities for teaching staff of educators in a design studio (with analytical, synthetical and metacognitive learning goals) is twofold: (1) they provide critique, constructive feedback and (2) they teach fellow students to criticize each other's approaches, to facilitate critical thinking in order to question their preferences and knowledge about the given problem (Gray, 2013). This pedagogical tradition aligns with theories about critical reflection involved in the constructivist pedagogical approach (Schön, 2017). To formulate critique and feedback in the form of questions has at least two advantages (Wiliam, 2013): questions implicitly cause thinking in students and student responses provide the teacher with information about the learner's current understanding in order to guide reflexive instructional practice. When teachers perceive student work samples and responses as feedback to themselves about the effectiveness of their teaching, they can indeed learn through critical reflection (Hattie, 2009). Even more than in traditional product design, the design students are challenged to cope with an increase in complexity, i.e., a multitude of interacting variables and stakeholders to consider. Hereto, the flexible use of clarifying representation and communication skills are key enablers to communicate clearly about the various design phases in order to receive constructive feedback to improve the PSS concept.

To exchange feedback between students and teachers, we used Comproved as a CJ–system, which was integrated into the university Learning Management System (LMS). For more information, see the Comproved practical guide for instructors (https://comproved.com/wp-content/uploads/2021/02/022021_Practical-guide-instructors.pdf) and the Comproved practical guide for participants (https://comproved.com/wp-content/uploads/2021/02/022021_practical-guide-participants.pdf).

## Theoretical Background on Feedback in Higher Education

### Comparative judgement

A growing body of literature supports the notion that comparative judgment can help learners and assessors in different learning and working situations (Lesterhuis et al., 2016; Van Gasse et al., 2017). CJ asks an assessor to compare two products and rank one product in relation to another. 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 for a real–life solution to a problem. In order to realize multiple comparisons using multiple judges (students, teachers and/or external assessors), a measurement scale can be created using the Bradley-Terry-Luce model (Bradley & Terry, 1952; Luce, 2005) showing the relative quality of each product (Pollitt, 2012). Furthermore, ideally each product receives 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 raters. Or its validity is anchored in what is valued by the community of practice within a given discipline (Jones et al., 2015).

Studies indicate CJ can be beneficial during a learning process (Bartholomew et al., 2019; Bouwer et al., 2018). CJ overcomes certain shortcomings of rubrics, which are often far too abstract for students to really grasp what 'quality' is, even those which specify performance levels and/or standards for each criterium, (Brookhart, 2018). CJ shows exemplars to students which are more relevant when it comes to understanding what quality is (Boud, 2000; Carless & Boud, 2018; Nicol & MacFarlane-Dick, 2006). When analyzing exemplars, students experience for example how high-quality products differ from average products (Orsmond et al., 2002).

Furthermore, learners are taught to reflect why product "A" is better than product "B", and learn to articulate why one product is better, worse or equal to another when they compare products. This thinking process is often referred to as metacognition, whereby the learner attempts to understand and explicitly name the intricate aspects which define why one product is better than another. When learners compare each other's PSS, both during and at the end of an educational process, learners also compare each other's analytical and synthetical thinking process, which might improve their metacognitive awareness of the product development process.

Another advantage in comparison to other types of assessment (e.g., rubrics) is that CJ focusses more on the holistic process and the holistic project result (Figure 1). Hereby the whole is more than the sum of its aspects. (Goossens & De Maeyer, 2018; Van Gasse et al., 2017). The PSS effort is directed towards a multilayered, complex challenge. A holistic assessment, building upon frequent formative feedback checkpoints is expected to result in a more valid grading for such complex assignments.

However, not all studies report overall or exclusively positive effects. Bartholomew et al. (2019) for example found that feedback given by students in a CJ setting can be rather superficial, e.g., mainly limited to the aesthetics of a certain design and rarely addressing more in-depth issues or providing more holistic feedback. Other studies have indicated that not just any kind of products can be compared. Comparisons of very extensive products with large quantities of information (such as >15 minutes movies), or products which are too different, (Bartholomew et al., 2019; Slovic & MacPhillamy, 1974) do not lead to reliable results.

**Types and levels of feedback**

Hattie and Timperley's (2007) Model of Feedback identifies different types and levels of feedback and considers the differing learning states of students. The different types of feedback are differentiated by the kind of question they answer. Feed Up (1) informs students about the goals, the learning intent, thus about: "where am I going?' Feed Back (2) informs them about "how am I going?", while Feed Forward informs them about "where to next?" However, not all types of feedback are considered equal. Gamlem and Smith (2013) state that students perceive feedback to be most effective when it includes improvement focused feedback that clarifies the next steps for learning. Boud and Molloy (2013) also emphasize that feeding forward should be an innate quality of feedback. Each of these feedback questions works at four feedback levels: task, process, self-regulation and the self-level.

At the task level, feedback is given about the specific requirements of the task, about how well the tasks are understood/performed. At the process level, feedback is directed towards the processes, skills, strategies and thinking required by the learner to understand and perform the task (the PSS design in our particular learning context). At the self-regulation level, the students are challenged to use deep learning principles such as relational thinking and self-monitoring to compare, direct, regulate and adjust their work in relation to the required standards (Butler & Winne, 1995). At the self-level, feedback focuses on personal, mostly positive evaluations of the learner, often associated with praise. In terms of implementation, Brooks et al. (2019, p. 27) indicate: "it is important to emphasize the ongoing interaction between the three feedback types rather than seeing them literally as boxes to be ticked off in linear fashion. Likewise, the

progression of feedback level is non-linear and relies on teachers' use of formative assessment practices to check their students' level of learning."

**Research Question**

As far as we know, no research into CJ has indicated what the quality is of peer- and teacher feedback and how this contributes to the design process and the acquisition of design competencies. As such this study seeks answers to the main research question: How can CJ become an effective formative assessment-tool to help students and lecturers provide clear formative feedback to students to improve their learning process during their PSS design project? As such, this study seeks to support the teachers to consider their impact and peer students' impact upon learning and to provide impetus to adjust instruction and future feedback processes during the assessment practice of the particular learning context of PSS design.

To gain insights into our main question, we formulate sub questions to gain further insights.

First, we explore (a) which types (and levels) of feedback are generated when applying CJ to guide the students' learning process and design project progress?

Second, we tentatively investigate (b) if and how the quality of the feedback given during CJ checkpoints during the PSS design process, can be further improved. Hereto we investigate:

(b1) which strengths and weaknesses do students experience with CJ-generated feedback?

(b2) which strengths and weaknesses do teachers experience with CJ-generated feedback?

(b3) which opportunities can be identified to improve the CJ-generated feedback?

Our study thus aims not only to explore which types and levels of feedback CJ based assessment with Comproved seems to foster. We also seek to better understand which are the perceived strengths and weaknesses by students and teachers of CJ with Comproved and which opportunities of improvement can be pinpointed for formative feedback during a PSS learning process.

## Methodological Approach

In this study, we describe a case-study of a Masters' design course with seventy-two master students that were placed into teams of four students. Each team had to collaborate -partly on-campus, partly online, due to the worldwide Covid-19 pandemic - on a semester-long PSS design project. Most students did not have previous experience with the PSS design toolkit, neither did they have experience in PSS design. To achieve the objectives for this twelve-week design course, student teams are expected to generate user insights and explore new opportunities to define and design a relevant PSS concept for this year's prompt 'The Future of Urban Health' and its proposed subthemes: (1) Your city, your vaccine (fit city); (2) The super responsive /resilient hospital; (3) The city without a hospital (micro hospitals).

Each week one third of the student teams has been asked to present their current project status in a short presentation at the start of the weekly consultation day, resulting in three checkpoints for formative assessment using CJ. Half-way in their design process, all student

teams present their intermediate project results in the form of a presentation during a mid-term jury. By the end of the project all teams once again present their project results during the final project presentation. Upon these three different project checkpoints, students, three aspirant teachers and four experienced design coaches provide formative feedback on the project progress. During the course we used a CJ tool, Comproved, to allow students to give feedback and rank each other's weekly, mid-term and final presentations. The required presentation form was a short video/movie (with a duration ranging from min. five to max. ten minutes) on their PSS project status: their design process, progress and intermediate project results, which allowed a comparison between the different student teams and their project achievements so far. In the middle and at the end of the design course, an assessment moment was organized to evaluate the projects. The evaluative role of CJ during the project process was both formative (for weekly consults, mid-term and final presentation) and summative (for the final presentation). The feedback, provided on the final assessment was both formative and summative because (a) feedback was formulated as such that it is applicable to future design projects or (b) when students do not pass in June, the feedback is applicable to improve their PSS design for resubmission in September.

In the following paragraphs, we will detail the relevance of a holistic assessment in the context of product development education. Table 1 on the next page offers an overview of the various checkpoint with feedback during the PSS project process in 2020 versus 2021, to show the impact of Covid-19.

***Table 1. an overview of the various checkpoints with feedback during the PSS project process in 2020 vs 2021***

| Feedback check-points | Weekly design day | | Mid-term assessment<br>Formative feedback for each team | Final assessment<br>Summative* feedback for each team |
|---|---|---|---|---|
| | Weekly consult timeslot to discuss project progress with 2 coaches<br><br>Formative feedback for each team | Presentations on project progress by 1/3 of all teams to the whole class<br>Formative feedback for 1 third of all teams each week | | |
| Pre-pandemic feedback from other students | Internal team member feedback; possibility to watch other teams' work on paper & to ask them for informal feedback on Campus | All teams can give feedback (and evaluate their own progress in comparison) after the presentations by 1/3rd of all teams on campus | All teams attend all presentations in an auditorium on campus, and are stimulated to ask questions & give feedback to all presentations | All teams attend all presentations in an auditorium on campus, and are stimulated to ask questions & give feedback to all presentations.<br>Students grade each other's expo posters, based upon 6 criteria |
| Pre-pandemic feedback from teachers | Oral feedback during +/- half an hour consults from a tandem of 2 coaches, live on Campus | Feedback by 4 teachers after each presentation | Grades - 6 grades for 6 criteria to access the PSS: Fit with nodes: Relevance; underpinning; logical build-up; elaboration; originality | Grades - 5 grades for 5 criteria to access the PSS:<br><br>Relevance; Verification & testing; logical build-up; elaboration and completeness; originality |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | - 2 grades for quality of presentation and quality of report | | - 3 grades for quality of presentation, expo poster and report | |
| In-Pandemic feedback from other students | Internal team feedback and possibility to watch other teams' work online on MURAL & to ask them for informal feedback online | 6 teams have to upload & rank each other's video & provide feedback to support each ranking on Comproved (evening before design day); the other teams can watch these videos | Written feedback based on Comproved Videos, guided by 4 criteria (on average 102 comments from 40 students for each team/project) | | Written feedback based on Comproved Videos, guided by 4 criteria (on average 102 comments from 40 students for each team/project) | |
| In-Pandemic feedback from teachers | Oral feedback during +/- half an hour consults from a tandem of 2 coaches, pre & post-lockdown: on campus, during lockdown: online + written feedback posted on team's canvases on MURAL | Oral feedback by 2 coaches (and 2 PhD students) during the online consults of the 6 teams | Written feedback based on Comproved Videos, guided by 4 criteria on average 8 comments from 3 experienced & 3 novice teachers) | Written feedback for subparts: prototyping quality, concept, video presentation, based on rubrics | Written feedback based on Comproved Videos, guided by 4 criteria (on average 8 comments from 3 experienced teachers & 3 novice teachers) | Written feedback for subparts: prototyping quality, concept, video presentation, based on rubrics |

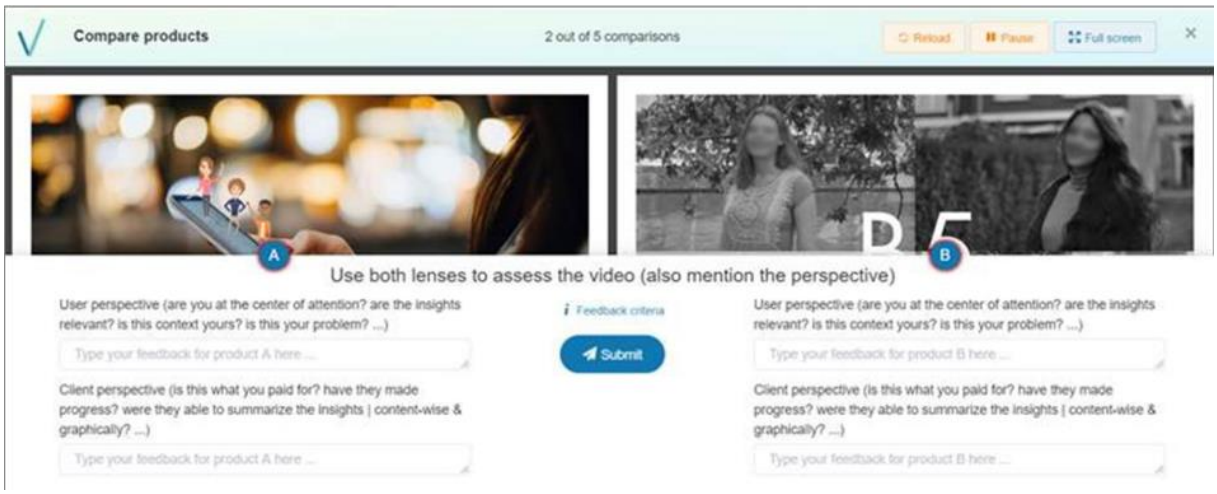


***Figure 1. An example of a comparative judgment checkpoint in Comproved, which asks students to indicate either A or B as the better product and provide feedback based on certain criteria.***

Students show their project progress online on Mural. Mural is a digital whiteboard for sharing and structuring unstructured information (e.g., pictures or audio files). It enables virtual collaboration on creative work, which can take place simultaneously or asynchronously (Lattemann et al., 2017) (For more details, see: https://www.mural.co/).

While pre-pandemic, the students received large A2 paper posters, printed with PSS tool templates to collaborate on, during the pandemic these templates were provided online to all

teams in Mural, as backbone to support the design process of students who had to collaborate online in a lock-down situation. Thus, the student teams could collaborate online on Mural Canvases with and without PSS tool templates in order to share their project progress and prepare their weekly consult moments, while teachers could track student teams' progress on Mural, leave post-it notes with feedback and suggestions and refer student teams to other team's Mural Canvases as referential exemplars for the right level of elaboration of certain tool applications or project progress expectations.

## Qualitative data gathering and data analysis

A qualitative study has been set up to better understand how CJ influences the students' learning process. In Comproved, we ask students to provide feedback on positive and negative aspects of both the PSS concept and the video, and strong and weak points of the PSS design, from a user and a client/investor perspective. The total amount of feedback on all projects from all teachers and students for the midterm checkpoint comprised 24,499 words, distributed over 2,749 feedback lines. A variety of qualitative data gathering techniques has been used in this study to obtain detailed information about the influence of CJ with Comproved on feedback types and levels. This approach permits data triangulation, as advocated by Yin (2016) to strengthen the credibility of a study by detecting convergence on outcomes. Two researchers independently screened the types and levels of feedback obtained using Comproved. Inconsistencies were discussed until consensus was reached.

Two researchers also divided the obtained students' and teachers' feedback about their CJ experience with Comproved in three broad categories: (1) perceived strengths, (2) perceived weaknesses of CJ, and (3) identified opportunities to improve the CJ-feedback quality. Afterward, feedback on which we did not reach consensus, was discussed and recategorized.

To answer the b1, b3 and b5 questions, two researchers have independently screened the students' reflections on their general learning experience (of a maximal length of half a A4 page) for feedback and comments from the students about their experiences and evaluation of the CJ-based feedback quality on Comproved. All students (n=72) submitted a reflection page. A qualitative analysis of the students' responses to an additional qualitative survey, with both broad and specific questions on educational quality (n=18) (Spooren et al., 2007) was also carried out, based on content analysis. This survey aimed to obtain more detailed insights into the educational experiences. To avoid bias, we did not ask specifically about Comproved, neither in the instructions with guiding questions for the reflection page, nor in the survey.

The teachers' feedback about their experiences and evaluation of the CJ-based feedback quality has been gathered by asking the four experienced design coaches to reply in writing to the questions b2, b4 and b5. The three apprentice teachers have not been included, as this semester was their first teaching experience. Therefore, they did only participate partly (providing the teams' short video pitches from feedback) instead of completely to the CJ evaluation & feedback process.

This article has tested a 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 (Laming, 2004) in which people tend to be better than in making absolute judgements (Thurstone, 1927).

## Findings and discussion

To gain insights into how CJ can become an effective means to provide clear formative feedback to students, we investigated which types (and levels) of feedback are generated to guide the students' learning process and design project progress when applying CJ. When commenting on the value of CJ-generated feedback, the students and teachers clearly indicate there are considerable differences in perceived value and quality of CJ-based feedback on Comproved between the individual students and between the student peer feedback versus the teacher feedback.

Therefore, we tentatively explore if and how the quality of the feedback given during checkpoints, can be further strengthened while using CJ. Hereto we have screened the feedback received from the students and the four experienced design coaches about their experiences and evaluation of the CJ-based quality of feedback. How do students perceive the influence of CJ on the feedback they give and receive during the PSS design process? Do students feel CJ is a valuable tool to give, receive and interpret feedback?

### Types and levels of feedback generated by CJ (Research question a)

The majority of feedback on the positive aspects is retrospective, thus more of the Feed Back Type, providing confirmation for certain choices made by the student teams. Also, in other studies (e.g., Brooks et al., 2019) Feeding Back was the most common type of feedback.

However, it is noteworthy that the comments on the negative aspects include more points of improvement. As such, they generate more interesting actionable, formative feedback of the Feed Up Type and also some of the Feed Forward Type for the students, which support the teams more to know which next steps to take, than Feed Back would do (Boud & Molloy, 2013; Gamlem & Smith, 2013). When we compare the levels of feedback, present in the feedback overview of Comproved, the task level is clearly dominant, with a second place for process level feedback. The self-regulatory is absent and the self-level scarcely appears amongst the feedback, generated by Comproved. These findings replicate studies on feedback in education, where firstly, process level feedback was consistently reported to be less frequently occurring to task level feedback (Brooks et al., 2019; Gan, 2011; Van den Bergh et al., 2013). Secondly, the 3 latter studies found feedback was directed to self-regulatory levels on only 1 to 2% of occasions relative to the other feedback levels.

However, we may also attribute a partly responsibility to the instructional design in the Comproved set-up, as students have only been asked to compare the videos and PSS concepts of other teams, not the video of their own team versus the video of another team. It might be

interesting to include these comparisons between own and other's work as well. However, one senior teacher (T1) expressed a genuine concern that this might lead to opportunistic and strategic ranking by students, who might rank the own project higher than any other, even if the other might be clearly superior to the own work. The guiding questions, which we implemented in Comproved, rather steer towards feedback on the task level, less on the process level and not at all on the self-regulatory and self-level.

Furthermore, a short instructional video on Comproved, pointing out the advantages of more Feed Up and Feed Forward Types of feedback, when possible, formulated in the form of questions (Wiliam, 2013), with equal attention to the tasks and process level might contribute to richer and more instructive, constructive feedback.

**Perceived strengths of CJ**

As experienced by students (Research question b1)

- CJ stimulates more peers to formulate & receive more feedback on each other's project status: "If we had given a live presentation in an auditorium for the mid-term, we would not have received as much feedback as we do now"
- The possibility to comment on the positive and negative aspects allowed students to obtain constructive feedback from peers on their work: "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 (c) and to prepare us for creating the final video(s) (d), and the other videos were also inspiring for our own work". "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 offers an overview of all the other products so students can explore them still on later occasions, to (a) see the progress and approach of other teams, (b) assess their own progress and presentation skills in comparison to others, and (c) asks feedback from others.
- CJ indirectly stimulates an overall better presentation quality of the students' products, by challenging them to present their project process in a self-explanatory 5 till 10 minutes video.
- The CJ software offers the ability to open the rankings so that both place (quality rating) and feedback on other work is available (see Figure 2). This is perceived as a great learning opportunity.


***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.***

As experienced by four experienced design coaches (T1>T 4) (Research question b2)

- CJ allows for different views of the results, both from students and teachers (T2), thus it results in better evaluations by and for all parties (T2). This may foster metacognition in terms of learning to communicate complexity and achieve simplification with the right representation.
- CJ stimulates a dialogue within the team and a competition between teams (T2). The benchmark between peers creates awareness in terms of better and worse, but also in judgement of best.
- Students can already proceed with the CJ feedback, even without their weekly consults (T2). In terms of envisioning possibilities and next steps (Feed Forward and Feed Up).
- Students generally don't want to offend anyone, out of some sense of justice (T1). Even so, it makes them think about giving feedback and gives them a better understanding of the design process).
- It objectifies your own assessment, compared to what students had ranked and the feedback they had given (T3). You can question yourself as teacher, offers a good benchmark.
- CJ requires students to review the results and express an opinion (T3).

**Perceived weaknesses of CJ**

As experienced by students (Research question b1)

- For interim products, student preferences go to the small-scale short presentation, a paper or a drawing. For large-scale products – describing full fledge solutions to a real-life problem (PSS design) – students prefer to compare and assess fewer products. They expect their peers would watch it more carefully and give more accurate feedback: “It felt like only a part of the video (product) was viewed each time, resulting in less informed reactions from people who did not completely understand, or gave comments

that were totally irrelevant, not useful, or at least much less than we (receiving the feedback) expected."

- Students also perceived a certain difficulty in interpreting the value of certain comments. For example, there is no hierarchy or difference made between the feedback on the product being from experienced design coaches, aspirant teachers or students. There is also no distinction made between the feedback of students who have dutifully seen the whole video, or who have only seen the first minutes. The teachers now compensate for this lack of differentiation between critiques, by discussing with the team the received CJ feedback, emphasizing priorities and separating the valuable from the confusing or irrelevant comments.
- CJ products must always be uploaded one day before the CJ, which results in a tighter deadline and more time pressure, as making a video takes much more time than e.g., preparing a presentation. Students also rightfully commented that there were no supportive courses to teach them how to produce a good video, leading to a much more time intensive learning-by-trial-and-error.

As experienced by four experienced design coaches (T1>T 4) (Research question b2)

- During CJ, a strategy of giving oneself a biased place in the group is always at play (T1). This is a concern that stands when giving scores, not when giving a rank and feedback.
- CJ obliges to give a ranking. Perhaps this is not justified. I don't see why students who have done well should necessarily still be distinguished from each other (T1). Referring to the 'excellence gap', differentiating up should not be an issue and provides feed-up for future design projects and as designer-professionals-in-the-making.
- A relatively high number of comparisons is necessary to reach a reliability over 0.7, (T2), e.g., difficult for intermediate (weekly) comparisons with only 1/3rd of the student teams, or when teachers divide the work (comparing lesser products each) for multiple deliverables.
- With CJ, it feels like the objective quality of learning goals / final terms is not assessed (T1). Results are less open for dialogue between teachers.
- You can only choose product A over B or vice versa and not judge A and B as equally good. It's not nuanced enough, it's more interesting to be able to say: I think a or b is better because there has been a lot of progress compared to last time, the story is right, this criterion scores higher and lower for A or for B (T3).
- CJ absorbs so much time, we should rethink the 'products' students upload to be shorter (T3).

Improvement opportunities for CJ -generated feedback (Research question b3)

- The instructional design can be improved. During the comparison, a clearly visible reminder or checklist to give better guidance about which feedback is expected for which criterium to obtain richer, more qualitative and qualitative feedback is necessary, best positioned near the feedback boxes. From following student's comment: "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" we learn that students experienced a dilemma about which aspects to focus on, when choosing the better 'product' and when considering which feedback to give. Thus, we find that some

students only base their comments on the video, others only about the underlying PSS concept, when they provide feedback from a user's and clients perspective, as shown in figure 1.

- A clear distinction should be made between CJ feedback coming from an experienced design coach versus from a meticulous or sloppy peer. However, information on who the feedback comes from and how much time has been taken by each particular "assessor" is available as the back end of the Comproved software, as shown in figure 3. Unfortunately, this info is not made available in an anonymous way to the student, nor to add a new layer to differentiate and rank feedback, according to which feedback is expected to be more trustworthy and reliable or less.

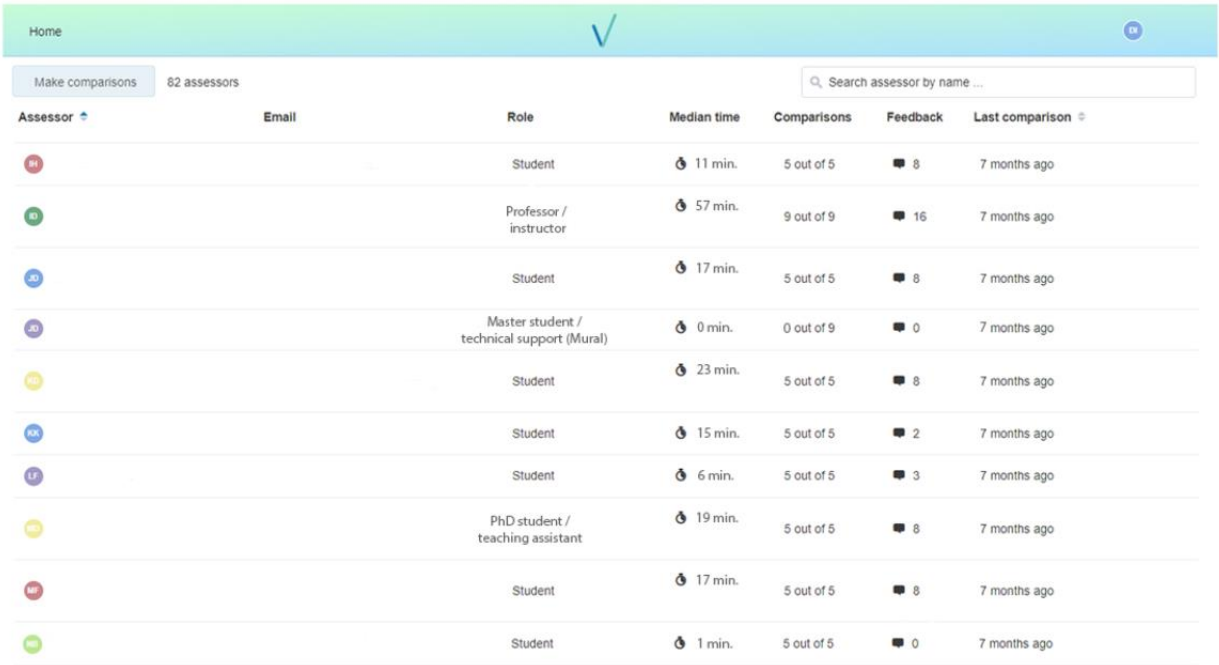


***Figure 3. The comparative judgment dashboard (only visible for teachers) offers an overview of the time spent and number of feedback comments given by each student.***

An indication credibility by Comproved could stimulate students to raise their feedback game and thus contribute to following point of improvement:

- More specific attention from teachers and peers, is indicated to raise feedback quality and variety in feedback types can support students in becoming more competent in providing constructive Feed Back, Feed Up and Feed Forward at the four levels. Providing more guidance and constructive Feed Back, Feed Up and Feed Forward at the four levels by means of regular coaching sessions can strengthen the further development of their pedagogical and design critique competences and design management skills.
- Specific timeslots should be reserved during the design exercise, so students and design coaches must reflect on their own and each other's feedback quality and underlying worldviews and biases, thus stimulating meta-cognition.
- Comproved should be more efficient in the follow-up of video-submissions and the processing of back-end information.

- Comproved should also provide the option of 'product' A is equal to 'product' B. Now you are forced to indicate that A is either better or worse, which is frustrating and seems unfair when both seem to be of the same quality.
- Students should be able to see a dashboard, to see how well they score on specific criteria, that keeps track of their evolution throughout the design process.
- As a teacher or student, it would be interesting to be able to benchmark and see how far you deviate from others.

## Conclusions

### Contributions of the study

This article describes which types of feedback Master students of Design and teachers formulate when they use CJ to provide feedback to each other, about the project progress during the design, and the final 'product' using CJ. Our findings show that the students perceived CJ to be helpful during and at the end of the design process, with certain reservations, as listed under "weaknesses". Furthermore we discern 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 are identified, supported by constructive Feed Up.
- A continuous comparison of intermediate 'products' supports student designers to make their doing, making and inventing explicit, which enables communication with others and provides a better understanding of the design process, which otherwise would remain largely tacit knowledge (even for the designer).
- CJ allows to streamline 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 by making the right representation choices. Comproved challenges them to produce self-explanatory, attractive videos to "sell" their PSS concepts to peers, future users and investors.
- 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.

We also suggest how CJ-generated formative feedback may be improved.

### Limitations of the study

Obviously, there are general limitations with respect to the fields of application of our results. In this paper, our focus is on using CJ, relying on the above-mentioned Comproved software with foremost formative feedback purposes, within an educational setting where students are designing PSS. We did not compare the results with other types of feedback systems. The additional qualitative survey in this study is also limited by its respondents (n=18). However, we argue that the students were overburdened at that time.

PSS and its complex representations require a holistic approach when it comes to feedback. COVID-19 brought new challenges to the design process and asked for additional skills to facilitate and guide online collaboration and meetings to receive and give mutual feedback amongst PSS development teams, multiple stakeholders, clients and users. We do not claim to

generate a representative sample, nor to generalize the results to other contexts, but rather to share our insights and opportunities to improve CJ-generated feedback in an online or hybrid teaching landscape.

**Future research**

A new hypothesis for future investigation thus emerged: might it be possible that a positive relation can be found between the quality (variety of types and levels) of feedback students provide to other student teams and the quality of their own PSS design results. In other words: do better design critics, providing more accurate and in-depth peer feedback, make better designers?

While CJ has proven to add to the understanding of the value of peer evaluation as part of students' active design education and reflection in this specific course, we still want to investigate more profoundly if and how CJ can contribute better to (indirectly) develop certain valuable skills and competences, necessary for our students, the design-professionals-in-the-making.

We would like to improve the formulation of instructions in Comproved, e.g., in the form of questions, to harvest more actionable feedback and steer more toward feed-up at a process level. Thus, we can evaluate if this leads students to formulate more qualitatively rich and reliable feedback of different types and different levels. As a possible next step for this research, it would be worthwhile to exchange experiences with design academics, who make use of CJ software to evaluate their students' design projects, with a comparable degree of complexity and duration (one semester) as the PSS project assignment. How do they seek continuous improvement in the way they provide feedback during a design project, evaluate the learning progress of their students and motivate them to become valuable co-evaluators of their peers?

## References

Bartholomew, S. R., Zhang, L., Garcia Bravo, E., & Strimel, G. J. (2019). A Tool for Formative Assessment and Learning in a Graphics Design Course: Adaptive Comparative Judgement. Design Journal. https://doi.org/10.1080/14606925.2018.1560876

Boud, D. (2000). Sustainable assessment for long term learning. Stud Contin Educ., 22(2).

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. Assessment and Evaluation in Higher Education, 38(6). https://doi.org/10.1080/02602938.2012.691462

Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing. Frontiers in Education, 3. https://doi.org/10.3389/feduc.2018.00086

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. Biometrika, 39(3/4). https://doi.org/10.2307/2334029

Brookhart, S. M. (2018). Appropriate Criteria: Key to Effective Rubrics. In Frontiers in Education (Vol. 3). Frontiers Media S.A. https://doi.org/10.3389/feduc.2018.00022

Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. Australian Journal of Teacher Education, 44(4). https://doi.org/10.14221/ajte.2018v44n4.2

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. Review of Educational Research, 65(3). https://doi.org/10.3102/00346543065003245

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. Metacognition and Learning, 11(2). https://doi.org/10.1007/s11409-015-9142-6

Carless, D. (2006). Differing perceptions in the feedback process. Studies in Higher Education, 31(2). https://doi.org/10.1080/03075070600572132

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. Assessment and Evaluation in Higher Education, 43(8). https://doi.org/10.1080/02602938.2018.1463354

Dewit, I. (2019). Product-service system design, a synthesis approach. University of Antwerp.

Dewit, I., Jacoby, A., & Matthyssens, P. (2021). Design Preconditions for Product—Service Integration. Designs, 5(2). https://doi.org/10.3390/designs5020029

Dewit, I., Van Ael, K., De Roeck, D., Baelus, C., De Rijck, R., & Coreynen, W. (2018). PSS Design and Strategic Rollout: tools for product-service systems (I. Dewit (ed.)). University Press Antwerp (UPA). https://www.aspeditions.be/nl-be/book/product-service-system-design-product-service-system-strategic-rollout/15791.htm

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. Assessment in Education: Principles, Policy and Practice, 20(2). https://doi.org/10.1080/0969594X.2012.749212

Gan, M. (2011). The effects of prompts and explicit coaching on peer feedback quality. The University of Auckland.

Goossens, M., & De Maeyer, S. (2018). How to obtain efficient high reliabilities in assessing texts: Rubrics vs comparative judgement. Communications in Computer and Information Science. https://doi.org/10.1007/978-3-319-97807-9_2

Gray, C. M. (2013). Informal peer critique and the negotiation of habitus in a design studio. Art, Design and Communication in Higher Education, 12(2). https://doi.org/10.1386/adch.12.2.195_1

Habermas, J. (1978). Knowledge and Human Interests (2nd ed.). Heinemann.

Hattie, J. (2009). Visible Learning: A synthesis of over 800 meta-analyses relating to achievement (1st ed.). Routledge.

Hattie, J., Gan, M., & Brooks, C. (2016). Instruction Based on Feedback. In R. E. Mayer & P. A. Alexander (Eds.), Handbook of Research on Learning and Instruction (2nd ed.). Routledge. https://doi.org/10.4324/9781315736419

Hattie, J., & Timperley, H. (2007). The power of feedback. In Review of Educational Research (Vol. 77, Issue 1). https://doi.org/10.3102/003465430298487

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing Mathematical Problem Solving Using Comparative Judgement. International Journal of Science and Mathematics Education, 13(1), 151–177. https://doi.org/10.1007/s10763-013-9497-6

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119(2). https://doi.org/10.1037/0033-2909.119.2.254

Laming, D. (2004). Human Judgment: The Eye of the Beholder. Thomson Learning.

Lattemann, C., Siemon, D., Dorawa, D., & Redlich, B. (2017). Digitization of the Design Thinking Process Solving Problems with Geographically Dispersed Teams. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and

Lecture Notes in Bioinformatics): Vol. 10288 LNCS (pp. 71–88). https://doi.org/10.1007/978-3-319-58634-2_6

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2016). Comparative judgement as a promising alternative to score competences. In Innovative Practices for Higher Education Assessment and Measurement. https://doi.org/10.4018/978-1-5225-0531-0.ch007

Luce, R. D. (2005). Individual Choice Behavior: A Theoretical Analysis (2nd ed.). Dover Publications.

Medola, F. O., Pavel, N., Baleotti, L. R., Santos, A. D. P., Ferrari, A. L. M., & Figliolia, A. C. F. (2021). Phenomenological Approach to Product Design Pedagogy: A Study on Students' Experiences in Interdisciplinary and Intercultural Settings. Design and Technology Education: An International Journal, 26(2), 86–100.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2). https://doi.org/10.1080/03075070600572090

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking Criteria in peer and Self-assessment. Assessment and Evaluation in Higher Education, 27(4). https://doi.org/10.1080/0260293022000001337

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. Assessment in Education: Principles, Policy and Practice, 19(3). https://doi.org/10.1080/0969594X.2012.665354

Puryear, J. S. (2015). Metacognition as a Moderator of Creative Ideation and Creative Production. Creativity Research Journal. https://doi.org/10.1080/10400419.2015.1087270

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences. https://doi.org/10.1007/BF01405730

Schön, D. A. (2017). The reflective practitioner: How professionals think in action. In The Reflective Practitioner: How Professionals Think in Action. https://doi.org/10.4324/9781315237473

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1). https://doi.org/10.3102/0034654307313795

Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. Organizational Behavior and Human Performance, 11(2). https://doi.org/10.1016/0030-5073(74)90013-0

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales. Assessment and Evaluation in Higher Education, 32(6). https://doi.org/10.1080/02602930601117191

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4). https://doi.org/10.1037/h0070288

Tovey, M. (2015). Design Pedagogy: Developments in Art and Design Education (1st ed.). Routledge.

Van den Bergh, L., Ros, A., & Beijaard, D. (2013). Teacher feedback during active learning: Current practices in primary schools. British Journal of Educational Psychology, 83(2). https://doi.org/10.1111/j.2044-8279.2012.02073.x

Van Gasse, R., Mortier, A., Goossens, M., Vanhoof, J., Petegem, P. Van, Vlerick, P., & Maeyer, S. De. (2017). Feedback opportunities of comparative judgement: An overview of possible

features and acceptance at different user levels. Communications in Computer and Information Science. https://doi.org/10.1007/978-3-319-57744-9_3

Wiliam, D. (2013). Assessment: The Bridge between Teaching and Learning. In Voices From the Middle (Vol. 21, Issue 2).

Yin, R. K. (2016). Qualitative Research from Start to Finish Second Edition. In The Guilford Press. https://doi.org/10.1007/s13398-014-0173-7.2