International Journal of Technology and Design Education
https://doi.org/10.1007/s10798-021-09712-3

Act like a user, work like a teacher: evaluation of experience design works through peer-testing

Armağan Karahanoğlu[1]

Accepted: 22 October 2021

Abstract
Experience design is a growing field that attracts many designers' attention. One of the challenges of experience design is to be emphatic with the users and be critical about evaluating the results of the design process. This challenge requires several skills and knowledge, which the designers could gain during their higher education. While user-centred design methods have extensively been used, teaching how to critically evaluate experience design from the users' perspective is barely addressed in design education. To address this challenge, I explored the suitability of peer assessment in two consecutive years of Design and Meaning, a third-year Industrial Design Engineering Bachelors' course at the University of Twente, the Netherlands. I followed a set of activities and developed peer-testing approach to amplify the role of design students in evaluating experience design works. In this approach, the students tested their peers' experience design works by pretending to be the users of the experience and assessed the works as if they were the experience design teachers. To prepare the students for this activity, I employed peer feedback and evaluation-preparation sessions during the course. Results showed that peer-testing could significantly contribute to experience design courses' evaluation-related learning goals when the students are provided with sufficient guidance. Results from both years provided promising evidence that peer-testing could be a method for teaching experience design evaluation in higher education design courses.

Keywords Peer assessment · Design education · Experience design · Peer-testing · Higher education

Introduction

It has been more than 25 years since experience design became the point of interest for many design research studies (Bongard-Blanchy & Bouchard, 2014; Hassenzahl, 2018). Experience design is defined as an approach in which the users' pragmatic and hedonic needs are positioned at the core of the design efforts (Hassenzahl, 2010; Hassenzahl et al., 2013). Hence, it

✉ Armağan Karahanoğlu
a.karahanoglu@utwente.nl

[1] Interaction Design Research Group, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands

requires understanding and reflecting on end-user needs, memories, and expectations (Desmet & Hekkert, 2007; Pucillo & Cascini, 2014). Researchers acknowledge that people are the experts of their own experiences (Gruen et al., 2002; Hassenzahl, 2010), and products that enhance people's experience increase products' success (Michalco et al., 2015). That is why experience designers advocate people's involvement in the design process from the early stages to understand, evaluate and reflect on experience diversity (Olsson, 2004; Pedgley et al., 2016; Wilkinson & De Angeli, 2014).

Acknowledging the importance of experience design, companies seek designers who develop interactive products that bring about positive experiences (G. Getto & Beecher, 2016). Therefore, skills for designing positive experience are expected to be in future designers' toolboxes. Meanwhile, several attempts have been made to integrate experience design topics into higher education programmes to prepare students for the job market (Gonzalez et al., 2017). However, equipping students with skills to identify the opportunities to design and test those opportunities in real cases requires knowledge in design and user-testing methods (Rose & Turner, 2020).

Educators in higher education seek ways to teach students how to evaluate the design process outcomes, both from the users' and the experts' perspectives (Yargın et al., 2018). In practice, researchers rely on a variety of user experience evaluation methods. These methods could both be qualitative (such as letting participants talk about their experience (Kujala et al., 2011)) and quantitative (such as using questionnaires and pre-defined measures (Effie Lai-Chong Law et al., 2014)). All these methods help the researchers arrive at rich evaluation results (Vermeeren et al., 2010). However, within the timespan of undergraduate courses, teaching how to evaluate the designed experiences becomes a challenge of higher education, as both planning and testing the experience design requires an extensive amount of time. Besides, due to the pace of the courses, it sometimes becomes difficult for students to carry out user tests with actual users. As a result, dealing with this aspect of experience design becomes one of the challenges of teaching user experience design (Guiseppe Getto et al., 2013). To address this, I came up with the idea of combining role-playing with peer assessment. The reason why I combined these two approaches is that role-playing is a technique that is employed in design process to empathise with users (Medler & Magerko, 2010; Simsarian, 2003; Svanaes & Seland, 2004), while peer assessment (Topping, 2009) is a well-established form of assessment in which peers assess other learners' products. Therefore, I hypothesized that peer assessment could be a suitable method to teach experience design evaluation in higher education. Hence, in this paper, I seek the answers to the question of *"What are the students' perceptions and learnings of employing peer assessment as a way of teaching experience design evaluation in higher education."*

To answer the research question, I redesigned the Design and Meaning course of the Industrial Design Engineering Bachelor's programme at the University of Twente. I first demonstrate background knowledge of teaching and testing experience design in higher education in the following lines. Following, I denote the course context and deployment of peer assessment as a form of experience design evaluation in higher education. In the end, I discuss the results of the two consecutive years and reflect on the future of peer assessment in design education.

Background

Design problems are ill-defined and require a certain level of expertise and experience to turn the problems into actionable design challenges (Cross, 2001, 2004). Löwgren and Stolterman (2004) advocate that designers are thoughtful about their role in the design outcomes and critically reflect on their decisions on people's lives and experiences. Reflecting on these roles makes the designers skilled to deal with complex and open problems (Dorst, 2011). This type of expertise requires a designer to intellectually use different design methods and tools in their career (Faiola, 2007).

In design, the view of "designing good looking products" was left many years ago. This view has been replaced with designing things that address fundamental (e.g. effectiveness, usefulness etc.) and affective (e.g. enjoyment, fun etc.) needs of their users (Hassenzahl, 2018), and that result in positive experiences (Effie L-C Law & Van Schaik, 2010). Giaccardi & Redström (2020) articulate that we are no longer living in a world that solely products exist, but more and more interactive "things" (e.g. products, data, AI, codes) do. The interplay between people and the ecosystem of these "things" yields experiences that are shaped by the interaction between them. One of the pioneers of experience design (Hassenzahl, 2013) states that experience design starts before designing products, and the experience becomes positive and meaningful when it fulfils people's psychological needs. One might think that addressing fundamental and psychological needs makes the design process even more complicated. However, this essentially highlights that human-centred design is at the core of good design practices.

de Bont & Liu (2017) argue that human-centred design can lead to breakthrough and design-led innovations and should be an integral part of design education. However, teaching experience design is intricate, because it positions in the intersection of multiple disciplines such as design, psychology and human–computer interaction (Rose & Turner, 2020). Therefore, design educators in higher education are becoming more interested in integrating human-centred design in teaching activities in educational programmes (Wormald, 2011). Such attempts aim to equip future designers to savvy the social context of the problem and design and test the solutions from theoretical and practical perspectives (Faiola, 2007).

In one of these attempts, Töre Yargın, et al. (2019) provide a detailed overview of different types of experience modelling that could be implemented in the early stages of experience design projects. They find that students model the experiences to make sense, communicate and act on them in the design process. In another one, Faiola & Matei (2010) emphasise the importance of developing students' critical thinking skills of understanding the impact of design on people's experiences. To achieve this, authors encourage students to use scenarios to envision how design can solve the complexity of interactive mobile devices' functions. Alternatively, Marti & van Leiden (2020) explore the inspirational effects of poems in the experience design process. They find that poems work like a means for students to extract more resonant qualities of experience to be translated into design properties.

Apart from designing the experiences, testing the design process outcomes (i.e. prototypes) helps understand how people experience the design process's real-life outcomes (Djamasb et al., 2016). Pettersson et al. (2018) discuss that it is sometimes difficult to understand what is being evaluated in user experience evaluations. This difficulty makes the improvement points challenging to extract from the outcomes. In practice, researchers define the criteria that the design should fulfil and design the peer-testing to assess

the design out of those criteria (Vermeeren et al., 2010). User testing can accumulate further knowledge about users and use-cases of designed experiences (Pagulayan et al., 2018). Therefore, designing and testing the experiences should be part of teaching experience design in higher education. However, even though there are examples of teaching experience design, testing the experience design in higher education received little attention. Hence, I turned to educational sciences to address this gap and to discover how design education can benefit from the learning assessment methods.

Employing peer assessment as experience design evaluation method

One of the higher education goals is to prepare students for the world's changing values while equipping them to be critical and independent evaluators of both their and others' works (Moesby, 2002). In that sense, student-centred learning is believed to attend to this goal by preparing students to keep pace with changing society (Hannafin & Land, 1997). Lee & Hannafin (2016) state that the "own it, learn it, share it" approach summarises the student-centred learning aims and goals. Accordingly, students play an active role in constructing and applying knowledge, while teachers guide them to own their learning goals and actions (Lee & Hannafin, 2016). With this approach, students grow into active learners and knowledge seekers independent of teachers.

There are multiple teaching and assessment methods in higher education that effectively facilitate student-centred learning. For instance, in-class activities allow students to have a voice in classes and increase students' engagement with the topics (Wright, 2011). This activity increases students' intrinsic motivation to participate in the activities (Ryan & Deci, 2000). Peer assessment has already been proven to be one of those in-class educational activities, which improves students' evaluation skills by being critical about others' work and their own (Lee & Hannafin, 2016).

Peer assessment has various positive effects on the students' learning process (Reinholz, 2016; Topping, 2009). Students feel more ownership of their learning process and are more responsible for it. They become more active in learning, pick up feedback better, and become more critical concerning the quality of their work and other people's work. The goal of the peer assessment is not to end up with a grade but to promote learning (Weimer, 2002). These are the effects aimed at student-driven learning, provided that peer assessment is appropriately used.

Peer assessment brings several benefits and challenges (Adachi et al., 2018; Boud et al., 2014). While it encourages students' ownership and responsibility of their learning (Topping, 2009), it also turns the students into active learners rather than being passive receivers of feedback (Adachi et al., 2018) and culminates positive effect on enhancing student's engagement (Kearney, 2013). Students can see various excellent and poor examples through peer assessment, enhancing their critical thinking (Hamer et al., 2015).

On the other hand, reliability and validity are the challenges of peer assessment. Several dynamics, such as friendship bonds, acceptance of criticism by peers, and power issues, can decrease peer assessment reliability (Zhang et al., 2020). Students' judgments and assessment of the work quality depend on their prior experiences and conceptions (Vu & Dall'Alba, 2007). The body of literature regarding students' and teachers' perspectives on peer assessment is relatively large (see for instance, Stigmar, 2016). Studies demonstrate that clear rubrics and structured assessment process, proper guidance, examples, and teacher guidance (Kearney, 2013; Topping, 2009; Van den Berg et al., 2006; Zhang et al.,

2020) can result in good or better assessments than teachers' assessments (Topping, 1998). Sivan (2000) also suggests being transparent during the process and discussing the benefits of peer assessment with students in advance.

Students can significantly benefit from the peer assessment approaches in learning how to evaluate the experience design works. As stated above, using peer assessment as a form of "user testing" is not reported yet. Leveraging this, I postulate that experience design evaluation can also benefit from the advantages of performance-based assessment (Linn et al., 1991). By definition, this type of assessment requires the students to 'showcase' what they learned rather than 'repeat' what they learned or memorised (McTighe & Ferrara, 1998). While the design students cope with the uncertainty of the design process, they can also take risks and the ownership of their learning (Seery et al., 2012; Wormald, 2011). Even though this type of assessment can be time-consuming (Dixson & Worrell, 2016), it fits the teaching experience design goals: evaluating the outcomes of experience design process.

In exploring the possible application of peer assessment, I recognised that peer-assessment in the form of role-playing could facilitate students' self-reflection on their design decisions. Role-playing has already been used as a design research methodology in teaching design in higher education, especially in design idea generation (Boess et al, 2007; Medler & Magerko, 2010). This method helps the design students to "accomplish in-depth reflection and discussion on the experience of interaction" (Boess et al., 2007, pg. 282). Therefore, swapping the roles with role-playing during the experience evaluation phase (i.e. taking user and experience design expert roles) could empower the students to behold how users could experience different designs.

Considering the critical role of assessment in the learning process, I propose peer assessment of experience design works can contribute to students' professional judgment in assessing an experience design's capability. Even though this approach is very similar to peer assessment, I merged this educational science terminology (i.e. *peer assessment*) with experience design terminology (i.e. *user-testing*). I came up with *peer-testing* as a better framing of our approach and expression of the educational activity I applied in the course. From this point forward, I will only use peer-testing to refer to the educational activity I employed in the course.

In the paper, I addressed two points: how does (1) peer-testing and (2) self-defined experience peer-testing criteria contribute to the learning goals of teaching experience design? There are two reasons why I specifically address these points. First, as I explained earlier, designers should be equipped with skills to understand users and be responsible for their design experiences' outcomes. Therefore, I expect the students to showcase the experiences they design and advocate the user needs while assessing others' experience design works. Second, every experience design can be unique (Karapanos et al., 2009), and therefore I would like to foster students' knowledge (i.e. "own it") by partially allowing them to define the peer-testing criteria of their experience design. In the following sections, I report the results of peer-testing activity in which I experimented in the Design and Meaning course, in 3rd year of the Industrial Design Engineering (IDE) program of the University of Twente.

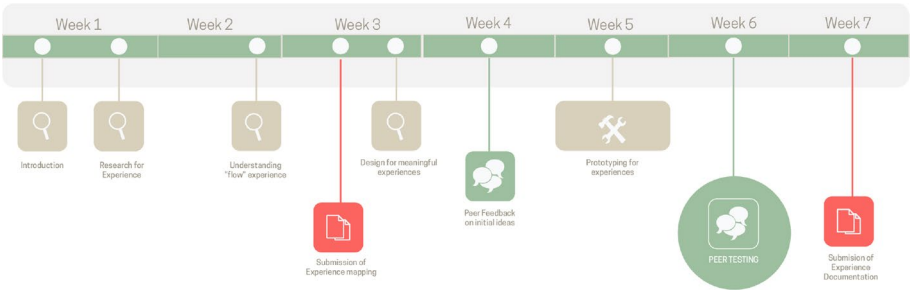

Fig. 1 The planning of the course

Methodology

To address the research question, I analysed the peer-testing results of the classes of 2019 and 2020.

Context of peer-testing

Design and Meaning course was offered in the eleventh quartile of the Industrial Design Engineering Bachelor's Education, the last quartile before the students work on their bachelor graduation assignments. Usually, a quartile length is ten weeks. However, the quartile in which the course is offered is shortened from ten weeks to seven weeks to open up extra time for bachelor's assignments. All students who follow the course have taken the program's mandatory courses, except their self-defined six months' minor studies. This makes me confident that students of the course possess the optimum knowledge and skills for completing the course.

The course is designed to get acquainted with several theories that unfold meaningfulness of people's experiences with interactive products. It aims to recall their human-centred design knowledge (taught primarily in the second year) to design interactive products that enhance people's experiences. Working in small groups of three or four, students first start mapping out people's experiences (Töre Yargın et al., 2019) and reveal their needs. Following, they design an interactive product to enrich people's experience, develop, prototype and test the experience. Finally, students report the experience design process in a visual document. Students spare 70 h (2.5 EC) individually in 7 weeks. Around 20 h are taken up by lectures, workshops and tutorials (Fig. 1).

During the lectures, students were introduced to theory around experience design, such as Experience Design (Hassenzahl, 2011), and psychology theory that is related to experience design, such as Self-Determination (Deci & Ryan, 2011) and Flow Theory (Csikszentmihalyi, 1990). The experience design assignment is formulated to spend the rest of the hours researching, designing, prototyping, and preparing the course's deliverables. The importance of attending the lectures, peer feedback and peer-testing sessions was emphasised at the beginning of the course. It was mandatory to be in the class during peer-testing sessions. In addition to the traditional teaching format (i.e. lectures), several teaching and peer-feedback sessions were utilised to support students' learning. The lectures were always followed by tutorials (1–2 h), in which guidance on how to integrate theory into the experience design was discussed.

Course goals, deliverables and assessment

After passing the course, students are expected to be able to:

- Recall the role of design at different levels of people's experience.
- Analyse people's experience of certain activities.
- Identify opportunities to design for people's experience.
- Design and test a concept by evaluating the models and frameworks of "design experience."
- Formulate report-like visual documentation of the experience design process

During the course, student effort was assessed based on three components of the assignment: (1) experience mapping (30%), (2) peer-testing (30%), and (3) experience documentation (40%). The teaching assistants and teachers do the grading of the first and third part of the assignment. The second part is entirely dedicated to peer-testing.

Students were provided with a rubric for different parts of the assignment. As was suggested by Asunda & Hill (2007), the contribution of each assessment criterium to the course's learning outcomes was indicated in the rubrics. We showed elsewhere that the rubrics we designed to assess the first and last part of the assignment were reliable and valid for assessing this course's learning outcomes (Karahanoglu et al., 2019). Therefore, I used the same rubric formats for the first and last parts of the assignment.

Design assignment

The assignment was to design an interactive physical product (and an app if that stands out from their research) to enrich people's physical activity experiences. The students were asked to form groups of three or four to work on the assignment. In the first week of the course, students negotiate and decide on the activity they would like to focus on. The assignment's focus was any activity that requires exertion, such as dancing, skiing, hiking, running, swimming, and cycling except daily activities (such as walking or cycling to work) as they were not interesting for the assignment. In the first weeks of the course, students explore how people experience their focus activity. They mapped out the selected activity experience and decided which type of interactive product would enrich the people's experience. In the end, the groups submitted a visual document (i.e. experience documentation) in which they explained the details of their experience of the design process and their final design.

Participants

I counted all the students as the participants of the study. In 2019 there were 70 students registered for the course, and in 2020 there were 86 students ($n = 156$ in total). In 2019, the students formed their working groups of 3 (2 groups) or 4 (16 groups) and summed up 18 groups in total. Like 2019, the students formed groups of 3 (10 groups) or 4 (14 groups), yielding 24 groups in total.

Table 1 Overview of peer assessment guidelines and matching peer-testing activities

Peer assessment guideline as listed in topping, 2009	Peer-testing activities applied in 2019 and 2020
1. Clarify the purpose and expectations of the assessment	I announced the aims and goals of peer-testing at the beginning of the course
2. Involve participants in defining the assessment criteria	The students were given the opportunity to define their experience design assessment criteria
3. Clarify the activities and timescale	Specified the activities and timescale before and during the peer-testing
4. Provide training, guidelines and checklists	*Provided training and examples of rubrics at the beginning of the course
5. Match and arrange contact between the peer groups before the assessment	*Announced the groups, matched the groups, arranged contact either in classroom or online
6. Monitor and coach	I monitored the peer-testing activity, coached the students where necessary
7. Examine the quality of the feedback	*I provided feedback on the quality of the feedback

*Some changes were made in 2020 either to ensure these guidelines or due to global pandemic

Activities of peer-testing

I planned the peer-testing moment in the sixth week of the course (see Fig. 1). During peer-testing, each group tested the experience design of other groups. To ensure the reliability and validity of the peer-testing, I followed the guidelines provided by Topping (2009). An overview of these guidelines and how I follow those guidelines are provided in Table 1. I explain those activities in detail in the following lines.

Guideline 1-Clarification of the purpose and expectations: Following the first guideline (in Table 1), I announced the purpose and the expectations of peer-testing in the first week of the course. I informed students that there would be several training moments to prepare them for this activity. Over the weeks, I reminded the students of the purpose and expectations of peer-testing as well.

Guideline 2-Involvement of participants in defining the assessment criteria: The students were given an overarching rubric that described the assessment criteria that applied to all groups. The overarching rubric consisted of three main criteria: (1) the value of the idea for the defined experience design purpose (10%), (2) design of the concept and its relation to the theory of experience (15%) and (3) the quality of prototype (25%).

In addition, groups were asked to define a set of assessment criteria (50%) based on which their experience design would be assessed (i.e. self-defined assessment criteria). I applied this approach as, in practice, researchers pre-defined a set of criteria that the experience design should fulfil during user peer-testing (Vermeeren et al., 2010). The idea here was that, even though the course has specific learning goals, each experience design can have unique qualities that should be considered in experience design peer-testing. Therefore, students were free to define the assessment criteria by considering their experience design goals and what they should achieve in the end. Both overarching and self-defined rubrics were utilised in peer-testing, and both took up an equal percentage in the final grade of groups. An example rubric is provided in the Appendix.

In 2019, students were not provided detailed guidance in setting up their self-defined assessment criteria. Instead, they were asked to define the criteria by considering their experience design goals and what their product should achieve in the end. However, this resulted in many questions from students, and hence in 2020, students were given a 20 min lecture on the self-defined assessment criteria and how to define them.

In both years, a minimum of five and a maximum of ten criteria was announced as a good number of self-defined assessment criteria. At the beginning of the course, students were provided with a template for writing down their experience design assessment criteria. In 2019, the student filled in the template printed and brought it to the peer-testing activity. Unlike 2019, in 2020, the groups defined and submitted the criteria before prototyping their experience design (Week 4). While defining proper assessment criteria, they were also asked to describe what they mean by the assessment criteria.

Guideline 3-Clarification of activities and timescale: At the beginning of each year, rules, activities and timescale of the activities of peer-testing were announced. In both years, students were announced that they would have 20 min to test the others' prototypes. During peer-testing, the designer group pitched their ideas and explained the experience goals to the assessor group. The assessor group was allowed to ask questions to comprehend the details of the experience design. After mutual peer-testing, each group had 15 min to decide on the group's grade and write feedback to the paired group. Afterwards, the groups came together, announced the assessment results and provided feedback to each other. I collected the filled rubrics at the end of the peer-testing activity.

Guideline 4-Guidance and training: To familiarise the students with qualities of an excellent peer-feedback session, I gradually introduced the idea of peer-testing over a few weeks before by,

- Grading the "experience documentation" with a rubric and providing detailed feedback over their work,
- Planning a peer-testing session during which the students provided feedback to the ideas of other groups,
- Guiding the development of self-defined assessment criteria.

A peer-feedback session was held in the fourth week of the course, during which each group gave to and received feedback from 3 different groups. The groups extended their idea with the received feedback, which they prototyped and peer-tested in the sixth week.

In peer-testing, each group had three roles: designer, user and teacher, which I announced at the beginning of the course. During peer-testing, the designer groups presented the prototype of their interactive product. The assessor group had then taken user and teacher roles. The assessor group first had the user role, in which they used the prototype as if they were the users of the designed experiences. The groups assessed the experience design and provided feedback as if they were the teachers of the course. After the peer-testing, the assessor groups were required to write detailed feedback about the experience design and provide evidence for the designer group's assessment results.

Guideline 5-Arrangement of contact: In 2019, the peer-testing activity was held in the classroom, with a capacity of 90 students. The students were already familiar with the classroom since all lectures of the course were given there. The classroom had movable

tables and other facilities for hanging papers on the walls. The students were asked to set their prototypes up in a spot they thought most convenient for peer-testing. Once the students were ready, I announced the paired groups, asked them to meet their paired groups and tested the designer group's prototype. In 2020, due to the worldwide Coronavirus outbreak, the peer-testing activity had to be held online and asynchronously. Unluckily, this was right before the students finalised their product prototypes. As soon as on-campus education was suspended, students were asked their opinion about the best possible way of explaining their experience design concepts via the university's online-class platform. In the end, it was mutually agreed that the groups submit a 2-min' video in which they explain their experience design, the ideal use case of their prototype and the way they think the product could enhance the experience. Students were given ten days to submit the videos to the university's online teaching platform (i.e. Canvas page).

Guideline 6-Monitoring: In 2019, the groups were matched and tested each other's prototypes. However, I had serious doubts that this might have affected the reliability of the test results. Hence, to increase diversity and ensure the reliability of test results, in 2020, each group gave and received feedback from 2 different groups. To achieve this, I shuffled the assessor groups so that none gave and received feedback from the same group. I announced this arrangement before the peer-testing activity.

In 2019, the peer-testing activity was possible to conduct on campus. I monitored the time and the contact of the groups. In 2020, the monitoring had to be online and asynchronous. I downloaded the videos and the assessment criteria files of the groups that each group submitted on Canvas. Via the Canvas page, each group was sent the works of 2 groups. The assessor groups had one week to assess the works and submit the peer assessment results on Canvas. I downloaded the results after the deadline and communicated the results of the peer-testing via Canvas. Due to the time limitation, student groups submitted the final part of the assignment (i.e. experience documentation) before receiving feedback from the assessor groups. This resulted in another limitation in that I could not compare the results of peer-testing with the grades of the experience documentation part of the assignment.

Guideline 7-Examining results: In 2019, the peer-testing was fully summative, and I only examined the results for my own understanding. However, in 2020, I decided to grade the quality of the given feedback. At the beginning of the class, the students were announced that the groups' feedback should be constructive rather than only summative. Therefore, I informed the students that I would assess the feedback provided to the designer groups. A rubric for the peer-feedback assessment was provided to the students before the peer assessment (See Appendix).

Data collection and analysis

In this paper, I address how the students perceive and learn from employing peer-testing as a way of teaching experience design evaluation in higher education. To this aim, I used data from three different sources.

First, I used the assessment criteria defined by students as a source of knowledge about the topics of the course. However, the wording and the appearance of the self-defined criteria varied among the groups. Therefore, I grouped the assessment criteria to overview each criterion's weighted effect on peer assessment's final grade.

Following,I analysed the assessment results of the peer-testing activity. First, I calculated the means and standard deviations of the assessment results of each part of the assignment and compared those with the peer-testing results. Next, I ran correlation analysis to investigate the relationship between the assessment results of designer groups and the assessment results of paired groups, and I ran repeated-measures t-tests to compare the results of peer-testing with the results of the final part of the assignment.

The final source of data was an online survey. After the final grades' announcement, students were asked to fill an online survey about their experience with the peer-testing activity. There were 3 question sets in the survey. The first set asked about the peer-testing grades, students' level of satisfaction of the grades and course in general, while the second set questioned how they think the peer assessment helped them achieve the course's learning goals. Those questions were rated from 1 (not at all) to 5 (very). Finally, students were asked to provide insights about how they think the peer assessment helped the course achieve the course's learning goals. In 2020, there were two additional question sets. The students were asked about the "fairness" of peer assessment and the assessment of feedback. In addition, students were asked their opinion about the effects of the Covid-19 global pandemic crisis on peer assessment. No other personal data, such as age and gender, was collected to ensure a safe space for students to provide feedback and to ensure anonymization of the results.

All the students were sent out the survey link via announcements and emails. They were requested to fill the survey within two weeks after the announcement of the grades. The survey collected anonymous data, and participation in the survey was voluntary. I downloaded the responses after the survey link was disabled (i.e., two weeks after the announcement). I calculated the mean, median and standard deviation of the responses of the learning goals related questions. To observe the similarities and differences between the years about students' perception of the achievement of the peer-testing activity, I ran independent samples t-tests. I grouped the responses to the open question of the survey to have an overview of students' perception of the effects of the peer-testing activity on their learning.

Results

In the following lines, I will first provide the numbers about the physical activity that the student groups focused on. Following, I will describe the results of assessment criteria defined by students, grades of peer assessment activity and responses to survey questions.

Focus physical activity of student groups

In 2019, the groups chose a variety of activities. These activities varied from individual activities (such as cycling, swimming and doing squat exercises), activities that people need at least one other person to perform (boxing and playing volleyball) to activities in which music is involved (such as drumming and dancing). In total, the groups addressed 13 different activities. In 2020, at the beginning of the course, students were given the list of activities that the class of 2019 selected. Surprisingly, in 2020, the activities the student groups selected for designing experience were entirely different from the class of 2019. Still, the activities included individual activities (such as bouldering and rope skipping), activities that people need at least one other person to perform (playing squash and

Table 2 The assessment criteria defined by student groups in 2019

Assessment criteria	Number of appearance (a)	Number of groups used this criterion (b)	Effect of the criteria on groups' grading (a/b)
Enhancement of experience	**21**	**16**	0.76
Interactivity	12	8	0.67
Intuitiveness	10	8	0.80
Feasibility of design	10	9	0.90
Usefulness	10	9	0.90
Ease of use	8	8	1.00
Design (not specified)*	7	5	0.71
Comfort in use	5	5	1.00
Functionality	4	3	0.75
Attractiveness	4	4	1.00
Effectiveness of use	3	3	1.00
Complexity*	2	2	1.00
Creativity*	2	2	1.00
User-friendliness	2	2	1.00
Personalisation	1	1	1.00
Pleasure in use	1	1	1.00
Total	102		

*Related to general aspects of design rather than the experience design

competitive rowing), as well as activities in which music is involved (such as pole fitness and dancing).

Results of self-defined assessment criteria

In 2019, 102 assessment criteria were defined by 18 groups ($M = 5.66$). The results showed that enhancement of experience appeared $n = 21$ times in the self-defined assessment criteria, while $n = 16$ groups listed these criteria in their rubric (Table 2). Interactivity was listed by $n = 8$ groups but appeared $n = 12$ times in the rubrics. Intuitiveness, usefulness, ease of use were listed by more than half of the groups. Other aspects of experience, such as pleasure in use, personalisation, creativity and user-friendliness, were explicitly listed by a small number of student groups. Interestingly, four of the self-defined criteria (highlighted by * in the Table 2) were about the general aspects (i.e., colours, forms) of the design or design process ($\pm 15\%$ in total). Still, all of the assessment criteria were relevant to the course topics (e.g. Flow Theory and Experience Design).

In 2020, 146 assessment criteria were defined by 24 groups ($M = 6.04$) (min 5, max 8). This average was higher than the class of 2019. The results showed that enhancement of experience appeared $n = 41$ times in the self-defined assessment criteria, while $n = 22$ groups listed these criteria in their rubric (Table 3). Following, intuitiveness was listed by $n = 15$ groups but appeared $n = 21$ times in the rubrics. The vast majority of groups listed feasibility of design, motivativeness and attractiveness. The results of 2020 also showed a relationship with the theory provided in the course (i.e. qualities of experience, usefulness,

Table 3 The assessment criteria defined by student groups in 2020

Assessment criteria	Number of appearance (a)	Number of groups used this criterion (b)	Effect of the criteria on groups' grading (a/b)
Enhancement of experience	41	22	0.54
Intuitiveness	21	15	0.71
Feasibility of design	11	11	1.00
Attractiveness	11	10	0.91
Motivativeness	9	6	0.67
Usefulness	7	6	0.86
Comfort in use	6	6	1.00
Engagement	6	3	0.50
Instructiveness	6	3	0.50
Interactivity	5	4	0.80
Inclusiveness	4	3	0.75
Complexity*	3	3	1.00
Ease of use	3	3	1.00
Personalisation	3	3	1.00
Design (not specified)*	3	2	0.67
Accuracy	2	2	1.00
Flexibility in use	1	1	1.00
Safety	1	1	1.00
Security	1	1	1.00
Sturdiness	1	1	1.00
User-friendliness	1	1	1.00
Total	146		

*Related to general aspects of design rather than the experience design,

Bold The criteria defined by the class of 2020 which is different from 2019

interactivity). Like 2019, three of the self-defined criteria (highlighted by * in Table 3) were about the general aspects (i.e., colours, forms). However, in 2019, fewer groups stated these in their self-defined criteria.

I found that in 2019, 9 groups (50% of the groups) had these vaguely defined criteria in their list, while in 2020, it was 5 (less than 25%). Thus, the number of criteria the class of 2020 defined was more than the class of 2019. This could be seen as a that students defined their self-defined assessment criteria was based on the details of their experience design.

In 2020, the groups clearly explained what they meant by the assessment criteria. Interestingly, some groups defined what "poor" and "excellent" in their rating scale mean. I received this evidence as a signal that the short lecture about developing self-defined assessment criteria helped the groups critically think about what they would like to achieve with their experience design.

Results of student grades

Results of students grades in 2019 showed that the range of overall grades of peer-testing activity falls between 6.65 and 9.35 ($M = 8.02$, $SD = 0.67$) out of 10. The average of the assessment out of the pre-defined criteria was $M = 8.00$ ($SD = 0.84$), while the average of

the assessment out of the self-defined criteria was $M = 8.03$ ($SD = 0.65$) out of 10. Results also showed a positive relationship between the assessment results of designer groups and the assessment results of paired groups, with a coefficient of $r = 0.51$, which is significant at $p < 0.05$. This shows that when the designer group's assessment results are high, the assessor group's assessment result is high. These pairs gave each other almost the same grade.

According to calculated average grades results, none of the groups failed the course in 2019 nor provided a supplementary assignment to pass the course ($M = 7.35$ out of 10). It should be noted that the passing grade is 5.50, and 6.50 is roughly estimated to be equal to "B" in the American grading system (Godor, 2017). The peer-testing results indicate that the peer assessment grades' average was higher than the accepted "good" grade of the Dutch universities (equal to 8.00). This was a questionable result, as it was not clear whether the experience design assignments were high in quality, or this result is due to the setup of the peer assessment activity. To interrogate this, I analysed the grades of the "experience document", in which the groups explained their final design.

In experience document part of the assignment, student groups explained the idea and theory behind their experience design. Groups further developed their design based on the received feedback. Results of 2019 showed that grading this part of the assignment fell between 5.00 and 9.00 ($M = 6.88$, $SD = 0.67$) out of 10. I compared these results with repeated measures t-tests. First, I checked the normality of residuals assumption with Q-Q, P-P plots and histograms. Test results showed that both scores are approximately normally distributed. Following I ran, repeated measures t-tests. Results showed that students' peer-testing grades were significantly higher than the grades of their final part of the assignment ($t(17) = 4.46$, $p < 0.05$, Cohen's $d = 1.35$). These results provided several impressions. The grades of peer-testing might have been high due to the activity's setup, that the groups gave feedback to each other and assessed their experience designs. Moreover, they communicated the results immediately after the peer-testing. Knowing this, I suspected that the friendship bonds might have influenced their peer assessment results.

In 2020, the range of the average grades that groups received from their peers fell between 5.20 and 8.47 ($M = 7.01$, $SD = 0.72$) out of 10. The average of the assessment out of the pre-defined criteria was $M = 7.02$ ($SD = 0.88$), while the average of the assessment out of the self-defined criteria was $M = 6.99$ ($SD = 0.68$). The grades of provided feedback (graded by teachers) were slightly higher than peer assessment results ($M = 7.40$, $SD = 1.21$). This shows that the quality of the given feedback was quite good.

Similarly to 2019, I also looked into the correlation between the grades groups received and gave. Results showed no relation between each group's received and given grades, with a coefficient of $r = -0.23$. This was not significant ($p > 0.05$). This result shows that the peer assessment grading of groups was not dependent on the grades they assessed. I received this result as a signal that the groups critically assessed their peers' experience designs.

Survey results

In 2019, $N = 18$ students (out of 70 students, 25, 71% of the class population) filled the survey. All of the participants indicated that their peer assessment grade was above 6.50 (out of 10). In addition, the results showed that the students were satisfied with the peer assessment activity grades *(*$M = 4.06$, $SD = 0.75$).

Table 4 Student evaluation of achievement of learning goals 2019 and 2020

Learning goal	Results of 2019			Results of 2020			*T*-test Results	
	Median	M	SD	Median	M	SD	*T*-score	Effect size (Cohen's d)
Recall the role of product design at different levels of people's experience	3.00	3.35	0.86	3.50	3.38	1.01	0.08	0.03
Analyse people's experience of certain activities	4.00	3.53	1.01	3.50	3.50	0.88	− 0.01	0.03
Identify opportunities to influence and design for people's experience	4.00	3.94	0.66	4.00	3.88	0.80	− 0.29	0.08
Design and test a product by evaluating the models and frameworks of "experience design"	3.00	3.41	0.71	3.00	3.21	1.06	− 0.73	0.22
Formulate design-based documentation of meaningful experiences	3.00	3.41	0.87	3.00	3.21	1.02	− 0.69	0.22

When I analysed the responses to the learning goals related questions (Table 4), the results of 2019 showed evidence that the most achieved goal was to "identify opportunities to influence and design for people's experience" ($M = 3.94$, $SD = 0.66$), while relatively least achieved goal one was to "recall the role of product design at different levels of people's experience." ($M = 3.35$, $SD = 0.86$). Overall, the median rating of the students was equal to or higher than the average value ($M > 3.00$) of the rating scale. In the end, students' overall satisfaction from the course was relatively high $M = 4.12$ ($SD = 0.70$).

In 2020, $n = 24$ students (out of 86 students, 27,91% of the class population) filled the survey. Of the participants, three indicated that their peer assessment grade fell below 6.50, while 21 indicated it was above 6.50. The results showed that the students were less satisfied with peer assessment results ($M = 3.54$, $SD = 0.88$) compared to 2019. The fairness of assessment of the assessor group was $M = 3.67$ ($SD = 1.01$), while the quality of the feedback received was $M = 3.54$ ($SD = 0.72$). Regarding the student's opinion about teachers' grading of the given feedback, three students indicated that their feedback grade fell below 6.50, while 21 indicated it fell above 6.50. The results indicated that they were more satisfied with the results of teachers' grading of the quality of given feedback ($M = 3.96$, $SD = 0.62$), the fairness of the assessment ($M = 4.33$, $SD = 0.48$), and the quality of the feedback received ($M = 4.25$, $SD = 0.61$).

The results also showed that with peer-testing activity (Table 4) the most achieved goal was to "identify opportunities to influence and design for people's experience." ($M = 3.88$, $SD = 0.80$), while the relatively least achieved goal was to "recall the role of product design at different levels of people's experience." ($M = 3.38$, $SD = 1.01$). Overall, the median of the students' rating was higher than the average value ($M > 3.00$) for all goals, indicating that students think that the peer assessment helped them achieve the course's learning goals. However, according to the survey results, the course's overall satisfaction dropped in 2020 ($M = 3.79$, $SD = 0.59$) compared to 2019 ($M = 4.12$, $SD = 0.70$).

An independent samples t-test was run for the learning outcomes part of the survey to compare the results of 2019 and 2020. I first checked the assumptions of independent samples *t*-test. The normality of the samples was checked again with Q-Q, P-P plots and histograms. Results showed that all scores are approximately normally distributed, while the sample sizes are not equal ($n = 18$ in 2019 and $n = 24$ in 2020). Levene's test for equality of variances for all questions are non-significant ($p > 0.05$ for all questions). Finally, I ran independent samples t-tests. Even though the results of 2020 were lower than 2019, there was no significant difference in the survey questions (see Table 4, the right column for T-scores of each question). The test results for the course's overall satisfaction were also nonsignificant ($p > 0.05$), $t(41) = 1.57$, $p > 0.05$, Cohen's $d = 0.51$).

Student feedback about peer-testing

In the online survey, students were asked to provide written feedback about their process of defining the self-defined assessment criteria. In 2019 some students indicated that they looked at their idea and defined the criteria by thinking about their design experience's unique achievements (6 students) or strengths (6 students). A few of them indicated that the assessment criteria they defined were important for their experience design (2 students) or the user group they targeted (2 students). One student indicated that they defined the criteria at the last minute by looking at what is possible to test. These results give the impression that in 2019, the students were already aiming at a high grade rather than thinking about their target experience design.

The survey results of 2020 showed that the groups of 17 students defined their self-defined assessment criteria by considering their experience design goals. Two respondents mentioned that they used their "experience" maps while defining their criteria, while two other respondents specifically mentioned that they used the "theory" provided in the lectures. One respondent stated that they set the criteria based on the feedback they received earlier from their peers. One stated that they were "free to set". One of the respondents did not state any reason.

In 2019, five students recommended more supervision about grading to improve the peer-testing process. This guidance was indicated to be on the detail of the self-defined assessment criteria and assessment process. Two students indicated that multiple groups should perform the assessment rather than only one, while one suggestion was that the teacher's assessment should be part of the peer assessment. The groups were less critical than the teacher, and they had to receive lower grades from the other group. One student indicated that they liked this process a lot.

The survey results showed that the students of 2020 were more knowledgeable about setting their self-defined assessment criteria. This is mentioned explicitly by one of the respondents: "*We tried to have a critical view of our own design*" (P03). The same participant also mentioned that the assessment criteria "*helped them improve the concept at an early stage*". Similarly, P12 mentioned that the assessment criteria '*have given structure to the last leg (i.e. prototyping) of the design process*', and P21 stated that '*The criteria were based on the goals that we identified for the product during and before the design process*'. All these results give the student groups the goals of their target experience design while setting their self-defined assessment criteria.

Effect of Covid-19 global pandemic crisis on peer-testing

Undoubtedly, in 2020, being asynchronous and online affected how the student groups tested each other's prototypes. This was also reflected in the survey results. Of the 24 respondents, 21 responded to the question of "How do you think that Covid-19 global pandemic crisis affected peer assessment results?". All the respondents stated that it negatively affected their experience. One student (P2) stated that '*With the Corona version, it was hard that you could not ask questions to the group you had to peer test*'. One of the stated reasons was that the students were not able to "touch" the prototypes, and this was a very important fallback of online peer-testing. Students indicated that touching and physically seeing is one of the most critical aspects of user-peer-testing and experience design. For instance, P23 stated that '*In general, I think I missed the interactive part of real-life peer peer-testing the most. You had to base your assessment entirely on only one video now*'. P7 stated that peer-testing over videos made the process "superficial":

> I think that because of the evaluation format the corona crises forced us to use, it was more challenging to understand the designs of other groups, making the feedback that we gave and received much more superficial.

Even though holding the peer-testing online and asynchronous stand out as a significant limitation of the process, the results showed promising potentials for online peer-testing experience design. I will reflect on this challenge and extrapolate the possibilities for applying peer-testing in online platforms in the discussions.

Discussion

Teaching experience design has become an interest of many educators in higher education (Wormald, 2011). Learning all the steps of experience design in higher education equips future designers with the skills to design and test their products from theoretical and practical perspectives (Faiola, 2007). Several studies investigate how to teach experience design in higher education (e.g. Faiola & Matei, 2010; Marti & van Leiden, 2020; Yargın et al., 2018). However, no study addressed the challenges of how to teach experience design testing in higher education. By looking into the educational sciences practices of assessment, I hypothesized that peer assessment could be a suitable method to teach experience design evaluation in higher education. I combined role-playing in user-*testing* (Medler & Magerko, 2010; Simsarian, 2003; Svanaes & Seland, 2004) with *peer assessment* (Topping, 2009) and employed *peer-testing* in two consecutive years (2019 and 2020) in [course name] course of IDE. Our goal was to teach design students to critically assess the design process's outcomes. To achieve this goal, I asked the research question of *"What are the students' perceptions and learnings of employing peer assessment as a way of teaching experience design evaluation in higher education?"*.

To adequately address the research question, I prepared the students for peer-testing. By following the guidelines of (Topping, 2009), I carried out a set of activities before, during and after peer-testing activity. I actively involved the students in the experience design evaluation process in facilitating student-centred learning (Sanchez-Elez et al., 2014). Before the peer-testing activity, students were given the freedom to select the type of experience they would like to design. The students were also given the freedom to define a rubric for their experience design evaluation. This was because, in practice, before user testing, design researchers pre-define the criteria that the experience design should fulfil (Vermeeren et al., 2010).

In peer-testing, students had three roles: designer, user and teacher. Undertaking the designer role first, in the first six weeks of the course, student groups designed and prototyped an interactive product to enhance the experience of a particular physical activity. During the peer-testing, students first undertook the user role. Then, they were invited to come together either in the classroom (in 2019) or online (asynchronously in 2020) to use and test the prototypes as if they were the users of the experience. Afterwards, they assessed the designer group's experience design by undertaking the teachers' role and provided feedback to the designer groups. With the feedback the groups received, they (were supposed to) further develop their experience designs. Thus, the entire course was designed to facilitate the peer-testing approach. The answer to the research question came from the analysis of terminology of self-defined criteria, the assessment results and an online student survey. I first employed peer-testing with the class of 2019, and where necessary, I made minor changes for the class of 2020 to improve the reliability and validity of the peer-testing results.

Reliability and validity of peer assessment became the major challenges of peer testing. As also stated in the literature, friendship bonds, acceptance of criticism by peers, and power issues decreased the reliability of peer assessment (Zhang et al., 2020). To tackle these challenges, I followed the guidelines of Topping (2009). Accordingly, I provided the students with clear rubrics and structured assessment process, examples, and teacher guidance (Kearney, 2013; Topping, 2009; Van den Berg et al., 2006; Zhang et al., 2020). I also prepared the students for this activity by discussing the benefits of peer assessment with students in advance as well (as was suggested by Sivan, 2000). Still, after the first

deployment in 2019, I had to make some changes in 2020 to address the reliability issues. For instance, in 2019, groups assessed and received feedback in pairs. I recognized that this resulted in notably high grades of the groups and I observed a correlation between the assessor and the designer groups' grades. To tackle this issue, in 2020, I changed the peer-testing setup: the groups gave feedback to and received feedback from two different groups. This change also helped investigate if peer-testing results in 2019 were high due to friendship bonds and power issues (Topping, 2009). The results of 2020 showed that this change resulted in receiving a relatively lower grade from the assessor groups. This result was evidenced in the student survey results as well.

Apart from the reliability issues, the results of the class of 2019 raised some doubts about whether this type of assessment is entirely suitable for teaching experience design testing in higher education. First, even though most of the self-defined criteria were aligned with the course content, some of the criteria seemed random (i.e. design, complexity, creativity) and not aligned with user peer-testing goals. Second, the grading of all the groups was significantly higher than the assessment of other parts of the assignment I graded, and the average of a good grade (i.e. 8 out of 10). Finally, in the online survey, students mentioned that while defining the self-assessment criteria, they thought about the design's unique achievements rather than what their aim was in the beginning. Therefore, I improved and reemployed peer-testing in the class of 2020 to see if peer-testing could turn into a learning activity for experience design courses (see 4.5 for the details).

It became evident in 2020 that self-defined criteria were aligned with the course content and even more detailed than in 2019. Besides, the randomness of the criteria was less than in 2019. I observed no correlation between the grading of the assessor and assessed groups. The survey results of 2020 showed that students thought about what they would like to achieve in their experience design and set their criteria accordingly while defining the self-assessment criteria. One of the drawbacks of letting the students self-define assessment criteria is that it could reduce the reliability of the assessment results. To resolve this challenge, I provided the group with an overarching rubric that each group had to fulfil the criteria, and hence the half of the partial grade of peer-testing came from this overarching rubric. I believe that self-defining the experience design assessment criteria is a unique activity of peer-testing and should not be avoided, as it mimics the pre-defining the criteria of real user-testing.

It was interesting that the peer assessment results were not correlated or aligned with the teacher assessment results in both years. It is already discussed in the literature that it should be not assumed that the teachers are more reliable than the students in terms of assessment (Topping, 2009). It also is dependent on the student groups' efforts in finalising the assignments. It might be the case that the students were more enthusiastic and engaged with the peer-testing (Kearney, 2013) than doing the other parts of the assignment. More often, it can be articulated that sometimes students love to design and construct something and put much energy into it, but they become less enthusiastic about reporting their process and findings.

In 2020, together with a teaching assistant, I also assessed students' written feedback. This was different from 2019, since I observed that the feedback students provided to their peers in 2019 was short in describing possibilities for improving the tested concepts. In 2020, teachers' grading of the written feedback consisted of half of the groups' grades of prototype testing activity grades. The survey results showed that the students were satisfied with their peer-testing grades and the given feedback. Knowing that the feedback would count towards a grade, students gave it more thought and effort.

While peer-testing can be regarded as a promising educational activity for teaching experience design evaluation, I observed several benefits and drawbacks of employing peer-testing. This observation was parallel to what educational literature reports. For example, I observed that peer-testing encouraged students' ownership and responsibility of their learning (Topping, 2009) which enhanced their critical thinking (Hamer et al., 2015). On the other hand, I also identified experience design-related challenges that design teachers should consider when employing peer-testing in their experience design course. That is, in real user-testing, the researchers encounter the real "experts" of the experience (Djamasbi et al., 2016), while in peer-testing, the students have to act like they are the users of the designed experience. This makes me question whether understanding the experience is sufficient to assess the 'enhancement of the experience' by role-playing the experience in the classroom. I acknowledge that this way of role-playing might stress the testing groups if they are not familiar with the activity they are in (e.g. can a non-swimmer understand the challenges of swimming experience?). To overcome this challenge earlier, I asked the student groups to model the experience in their earlier assignment (e.g. experience maps). Students can show their experience maps to the user groups during peer-testing. Still, I think the better way to solve this issue would be to match the student groups earlier and ask the designer groups to design for the experience that the matched groups are familiar with. This way, students' knowledge of both designing and evaluating the experience design would be stimulated.

As discussed above, the results from 2019 and 2020 showed that peer-testing has several potentials as a teaching design experience evaluation. Based on my experience and results of the studies, I provide six suggestions to future design educators in applying peer-testing in the following lines.

- *Introduce the goals and importance of peer-testing* The results I collected from 2019 evidenced that it might be difficult for them to understand why they assess one others' design outcomes while the teacher can assess the works. Peer-testing would be most useful when the students take the roles of both users and expert designers. Therefore, an educator who would like to employ peer-testing in design education should introduce why peer-testing is part of their teaching. If the roles the students could undertake are not essential or straightforward for the experience design courses, alternative methods to peer-testing should be considered.
- *Prepare the students for peer-testing* The literature and the results of the studies showed that preparing the students for peer-testing is highly important. This will give the students time and will reflect on what to expect from the peer-testing. Rubrics can achieve this while arranging a prior peer-feedback session would greatly benefit the students for peer-testing. In addition, the results showed that peer-testing should not be in pairs but a chain, to make the peer-testing more reliable and constructive.
- *Guide the students for choosing the activity to enrich* One of the limitations of testing the experience design with peers is that the user groups should be familiar with the activities that the designer groups design for. Hence, to reduce the stress over both the designer and the user groups, students could be guided about the more appropriate activities for the peer-testing activity.
- *Guide the students for defining self-defined assessment criteria* In practice, for testing the design outcomes, a group of researchers define the criteria that the design should fulfil (Vermeeren et al., 2010). Therefore, defining self-defined assessment criteria could be regarded as planning for experience design evaluation. However, students

must be guided about defining those criteria, how they should be prioritised and what it would mean for the users; in other words, the assessor groups.

- *Assessment of feedback increases the quality of feedback, while it is time-consuming for the teacher* After applying peer-testing in 2019, I decided to assess the feedback given by groups. This was done to encourage the students to give constructive feedback, be critical and fair in peer assessment. On the other hand, the assessment of feedback is time-consuming for the teacher, as it requires the teacher to look back and indirectly assess the experience design of the groups. Therefore, the teachers might consider not assessing the feedback if there is no assessment involved in the peer-testing activity.
- *Peer-testing should both be summative and formative* Based on the results of both years, it can be articulated that peer-testing should be summative and formative. In other words, it will be more beneficial for the students if the feedback received from their peers is used in the next step of the design process. This was also evidenced in mid-term peer-feedback sessions.

Conclusions

In this paper, I explained the necessity and challenges of teaching experience design in higher education. I addressed one of those challenges: teaching experience design assessment by developing and applying the peer-testing method. I applied this method in the Design and Meaning course of Industrial Design Engineering bachelors' programme in the University of Twente in two consecutive years (2019 and 2020). The first application in 2019 resulted in several hurdles, which were later addressed in 2020. Even though the global Covid-19 pandemic influenced the peer-testing deployment in 2020, the results presented in this paper provide evidence on the benefits of employing peer-testing in experience design evaluation in higher education.

I pointed out several challenges of employing peer-testing and ways to resolve them. One of the limitations of the study is that the end users never tested the designed experiences. Future research is required to test whether peer-testing results are aligned with actual user-testing results. Still, the main contribution with this paper is demonstrating how the peer-testing activity could be performed and turned into a learning moment for students of experience design classes in higher education. I encourage the design educators to use this method by following the guidelines provided in the literature and suggestions I described above.

Appendix

Rubric and guidance provided to the students for peer-testing

Overarching assessment criteria for peer-testing

End Result will be in between 0–1.5 points.

The owner of the prototype /design (Write the assessed group's number here): ……

Above mentioned group is assessed by group:……

The assessment of each criteria can be between the defined intervals, such as 0.3 / 0.7/ 1.2 / 1.3 etc.

Skills	Poor (0.5)	Average (1.0)	Excellent (1.5)	Feedback / Improvement points
Ideation	*The idea is not creative at all *It doesn't add anything to the experience of motion *Unclear user group	*The idea is somewhat creative * It adds minor qualities to the existing experience * Clear user group	*Adds surprisingly new approaches to the experience *Surprisingly interesting and creative idea * Clear user group	
Design	*Touches very little-to not at all- on Flow Experience, by showing limited knowledge of experience design	*Touches on the Flow Experience by showing basics of the theory, but lacks the major relevancies	*Makes an outstanding relevance with the Flow Theory, takes all dimensions into consideration and even goes beyond those.	
Prototyping	* Very poor prototype in terms of functionality and aesthetics * Almost not working * Does not show the idea / experience	* Working properly * Good in terms of quality of functions and aesthetics * Shows the details of the idea / experience * Shows all of the important aspects of experience	*Excellently working, * Shows excellence in quality of prototype * Excellently showing the details of the idea/experience; * Outstandingly shows all of the important aspects of experience	

Template for students to design their own assessment criteria for peer-testing

Fill the first column of the below table with the criteria that you want your prototype / design to be assessed out of and use the second column for what you mean by the criteria.

The assessment of each criteria can be between the defined intervals, such as 0.3 / 0.7/ 1.2 / 1.3 etc.

*The assessment criteria**	Explanation of Criteria	Poor (0.5)	Average (1.0)	Excellent (1.5)	Feedback / Improvement points
1.		Poor	Average	Excellent	
2.		Poor	Average	Excellent	
3.		Poor	Average	Excellent	
4.		Poor	Average	Excellent	
5.		Poor	Average	Excellent	
6.		Poor	Average	Excellent	
7.		Poor	Average	Excellent	
8.		Poor	Average	Excellent	
9.		Poor	Average	Excellent	
10.		Poor	Average	Excellent	

**Listing 5 criteria is preferable. This should not exceed 10.*

Assessment criteria for quality of feedback[1]

End Result will be in between 0-1.5 points.

Qualities of good feedback	Grade
Be **specific**. For example if information is missing, then it helps to state what is missing and where exactly. It also helps to give examples in the student's work. Avoid: "You really should substantiate your argument with more information." But say: "This and that specific information is missing in your piece."	
Try to give feedback only on **the work and the progress under discussion** and not on the student personally. This may sound fairly logical, but the division line is not always apparent. Avoid: "You haven't handled this very smartly." (has the student thinking: he thinks I'm stupid) But say: "You have not followed this or that guideline."	
Give feedback on what you can **concretely** observe in the student's work and progress and not on any interpretations thereof ('I' message). Everyone has an idea about why a student has approached or performed things in a certain manner, but you don't know this for certain and therefore it is better to avoid giving feedback on this. What you can do is talk to the student about why he/she does certain things.	
Ensure you also mention what the student has done **right.**	
Provide **feedback** (where you are now), as well as **feed up** (where you are going) and **feed forward** (what is your next step).	

Fundinag The project was sponsored and supported by the Center for Engineering Education (CEE) of the 4TU Federation in the Netherlands and by the University of Twente. More information can be found here: https://www.4tu.nl/cee/.

References

Adachi, C., Tai, J.H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*(2), 294–306.

Asunda, P. A., & Hill, R. B. (2007). Critical features of engineering design in technology education. *Journal of Industrial Teacher Education, 44*(1), 25–48.

Boess, S., Saakes, D., & Hummels, C. (2007). *When is role playing really experiential? Case studies.* Paper presented at the Proceedings of the 1st international conference on Tangible and embedded interaction.

Bongard-Blanchy, K., & Bouchard, C. (2014). Dimensions of user experience-from the product design perspective.

Boud, D., Cohen, R., & Sampson, J. (2014). *Peer learning in higher education: Learning from and with each other*. Routledge.

Cross, N. (2001). Designerly ways of knowing: Design discipline versus design science. *Design Issues, 17*(3), 49–55.

Cross, N. (2004). Expertise in design: An overview. *Design Studies, 25*(5), 427–441.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

de Bont, C., & Liu, S. X. (2017). Breakthrough innovation through design education: Perspectives of design-led innovators. *Design Issues, 33*(2), 18–30.

Deci, E. L., & Ryan, R. M. (2011). Self-determination theory. *Handbook of Theories of Social Psychology, 1*(2011), 416–433.

Desmet, P., & Hekkert, P. (2007). Framework of product experience. *International Journal of Design, 1*(1), 57–66.

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice, 55*(2), 153–159.

[1] Reference: https://www.ru.nl/lecturers/education/teaching-guiding/giving-feedback/constructive-feedback/

Djamasbi, S., Strong, D., Wilson, E. V., & Ruiz, C. (2016). *Designing and testing user-centric systems with both user experience and design science research principles*. Paper presented at the Twenty-second Americas Conference on Information Systems, San Diego.

Dorst, K. (2011). The core of 'design thinking'and its application. *Design Studies, 32*(6), 521–532.

Faiola, A. (2007). The design enterprise: Rethinking the HCI education paradigm. *Design Issues, 23*(3), 30–45.

Faiola, A., & Matei, S. A. (2010). Enhancing human–computer interaction design education: Teaching affordance design for emerging mobile devices. *International Journal of Technology and Design Education, 20*(3), 239–254.

Getto, G., Potts, L., Salvo, M. J., & Gossett, K. (2013). *Teaching UX: Designing programs to train the next generation of UX experts.* Paper presented at the Proceedings of the 31st ACM international conference on Design of communication.

Getto, G., & Beecher, F. (2016). Toward a model of UX education: Training UX designers within the academy. *IEEE Transactions on Professional Communication, 59*(2), 153–164. https://doi.org/10.1109/TPC.2016.2561139

Giaccardi, E., & Redström, J. (2020). Technology and More-Than-Human Design. *Design. Issues, 36*(4), 33–44. https://doi.org/10.1162/desi_a_00612

Godor, B. P. (2017). Revisiting differential grading standards anno 2014: An exploration in Dutch higher education. *Assessment & Evaluation in Higher Education, 42*(4), 596–606.

Gonzalez, C. A., Smith, M. A., & Youmans, R. J. (2017). *Are human factors students prepared for careers in user experience research? A survey of predicted and actual skill utilization.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Gruen, D., Rauch, T., Redpath, S., & Ruettinger, S. (2002). The use of stories in user experience design. *International Journal of Human-Computer Interaction, 14*(3–4), 503–534. https://doi.org/10.1080/10447318.2002.9669132

Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education, 40*(1), 151–164.

Hannafin, M. J., & Land, S. M. (1997). The foundations and assumptions of technology-enhanced student-centered learning environments. *Instructional Science, 25*(3), 167–202. https://doi.org/10.1023/a:1002997414652

Hassenzahl, M. (2011). *User experience and experience design*. The Interaction-Design.org Foundation.

Hassenzahl, M. (2013). User experience and experience design. *The encyclopedia of human-computer interaction, 2*.

Hassenzahl, M., Eckoldt, K., Diefenbach, S., Laschke, M., Len, E., & Kim, J. (2013). Designing moments of meaning and pleasure. Experience design and happiness. *International Journal of Design, 7*(3).

Hassenzahl, M. (2018). The thing and I: understanding the relationship between user and product. In *Funology 2* (pp. 301–313): Springer.

Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. *Synthesis Lectures on Human-Centered Informatics, 3*(1), 1–95.

Karahanoglu, A., Alink, C. O., & Bakirlioglu, Y. (2019). Quantifying design for user experience assignments using rubrics as assessment tools. Paper presented at the DRS learnXdesign 5th international conference for design education researchers 2019.

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.-B. (2009). *User experience over time: an initial framework*. Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA.

Kearney, S. (2013). Improving engagement: The use of 'Authentic self-and peer-assessment for learning'to enhance the student learning experience. *Assessment & Evaluation in Higher Education, 38*(7), 875–891.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473–483. https://doi.org/10.1016/j.intcom.2011.06.005

Law, E.L.-C., & Van Schaik, P. (2010). Modelling user experience–An agenda for research and practice. *Interacting with Computers, 22*(5), 313–322.

Law, E.L.-C., Van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies, 72*(6), 526–541.

Lee, E., & Hannafin, M. J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational Technology Research and Development, 64*(4), 707–734.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

Löwgren, J., & Stolterman, E. (2004). *Thoughtful interaction design: A design perspective on information technology*. Mit Press.
Marti, P., & van Leiden, F. (2020). Teaching experience design using poems as cultural probes. In *Project and Design Literacy as Cornerstones of Smart Education* (pp. 21–34): Springer.
McTighe, J., & Ferrara, S. (1998). *Assessing learning in the classroom. Student assessment series*. ERIC.
Medler, B., & Magerko, B. (2010). *The implications of improvisational acting and role-playing on design methodologies.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Michalco, J., Simonsen, J. G., & Hornbæk, K. (2015). An exploration of the relation between expectations and user experience. *International Journal of Human-Computer Interaction, 31*(9), 603–617. https://doi.org/10.1080/10447318.2015.1065696
Moesby, E. (2002). From pupil to student–a challenge for universities: An example of a PBL study programme. *Global J. of Engng. Educ, 6*(2), 145–152.
Olsson, E. (2004). What active users and designers contribute in the design process. *Interacting with Computers, 16*(2), 377–401. https://doi.org/10.1016/j.intcom.2004.01.001
Pagulayan, R. J., Steury, K. R., Fulton, B., & Romero, R. L. (2018). Designing for fun: user-testing case studies. In M. Blythe & A. Monk (Eds.), *Funology 2: From usability to enjoyment* (pp. 419–433). Springer International Publishing.
Pedgley, O., Rognoli, V., & Karana, E. (2016). Materials experience as a foundation for materials and design education. *International Journal of Technology and Design Education, 26*(4), 613–630.
Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., & Butz, A. (2018). *A Bermuda Triangle? A Review of Method Application and Triangulation in User Experience Evaluation.* Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
Pucillo, F., & Cascini, G. (2014). A framework for user experience, needs and affordances. *Design Studies, 35*(2), 160–179.
Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education, 41*(2), 301–315.
Rose, E. J., & Turner, H. N. (2020). *The Pedagogy of User Experience: Methods, Tools, and Approaches*. Paper presented at the Proceedings of the 38th ACM International Conference on Design of Communication, Denton, TX, USA. https://doi.org/10.1145/3380851.3416787.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68.
Sanchez-Elez, M., Pardines, I., Garcia, P., Miñana, G., Roman, S., Sanchez, M., & Risco, J. L. (2014). Enhancing students' learning process through self-generated tests. *Journal of Science Education and Technology, 23*(1), 15–25.
Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education, 22*(2), 205–226.
Simsarian, K. T. (2003). *Take it to the next stage: the roles of role playing in the design process.* Paper presented at the CHI'03 extended abstracts on Human factors in computing systems.
Sivan, A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education: Principles, Policy & Practice, 7*(2), 193–213.
Stigmar, M. (2016). Peer-to-peer teaching in higher education: A critical literature review. *Mentoring & Tutoring: Partnership in Learning, 24*(2), 124–136.
Svanaes, D., & Seland, G. (2004). *Putting the users center stage: role playing and low-fi prototyping enable end users to design mobile systems.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.
Topping, K. (2009). Peer Assessment. *Theory into Practice, 48*(1), 20–27.
Töre Yargın, G., Günay, A., & Süner-Pla-Cerda, S. (2019). *UX modelling in design education: methods, processes and examples*.
Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(03), 341–356.
Vermeeren, A. P., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). *User experience evaluation methods: current state and development needs.* Paper presented at the Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries.
Vu, T. T., & Dall'Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education, 32*(5), 541–556.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. John Wiley & Sons.
Wilkinson, C. R., & De Angeli, A. (2014). Applying user centred and participatory design approaches to commercial product development. *Design Studies, 35*(6), 614–631.
Wormald, P. W. (2011). Positioning industrial design students to operate at the 'fuzzy front end': Investigating a new arena of university design education. *International Journal of Technology and Design Education, 21*(4), 425–447.
Wright, G. B. (2011). Student-centered learning in higher education. *International Journal of Teaching and Learning in Higher Education, 23*(1), 92–97.
Yargın, G. T., Süner, S., & Günay, A. (2018). *Modelling user experience: Integrating user experience research into design education.* Paper presented at the International Conferences Interfaces and Human Computer Interaction 2018.
Zhang, F., Schunn, C., Li, W., & Long, M. (2020). Changes in the reliability and validity of peer assessment across the college years. *Assessment & evaluation in higher education*, 1–15.