Quantifying Design for User Experience Assignments: Using Rubrics as Assessment Tools

Armağan KARAHANOĞLU
University of Twente
a.karahanoglu@utwente.nl

Charlotte OUDE ALINK
University of Twente
c.oudealink@utwente.nl

Yekta BAKIRLIOĞLU
University of Limerick
yekta.bakirlioglu@ul.ie


Abstract: There is an increasing interest in teaching user experience design in many of the industrial design bachelor’s programs. The subjectivity of the topic requires new approaches as well as reliable and valid assessment tools. It has always been a challenge for the teachers to assess creative work in higher education. In relation, the assessment of how products create user experience in student works requires extra attention. In this paper, we discuss the difficulty of properly assessing design and explain the development and application of rubrics that we aimed to facilitate the assessment of design for user experience assignments of a 3rd year bachelors’ course of the University of Twente. We present evidence of the reliability and validity of the assessment through the rubrics. Usability of the rubrics for assessment purposes has also been addressed.

Keywords: assessment; evaluation; grading; design assignments; user experience assignment

1 Introduction

It has been more than 25 years that the term user experience (UX) has become a breakthrough in human-computer interaction studies (Hassenzahl, 2018). Since then, it has been regarded as an interdisciplinary field that studies human behaviour systematically to design useful and desirable products. Several frameworks have been developed to unfold the dimensions of people’s experience with interactive products and systems. Understanding the importance of the involvement of people’s experience in the design process, companies recently expect from designers a new type of expertise, so-called designing for user experience. However, according to our knowledge, there is no bachelor program that directly trains the students as user experience designers, and thus the ongoing discussion in filling this gap in design education is very significant. Several bachelors and master’s programs are designed to prepare students to the job market with skills to design for user experience. Even scholars discuss the importance of integrating UX into human-centred design related higher education programs (Faiola, 2007). Within this context, there are several efforts...
to generate an educational agenda for teaching user experience in higher education (Getto & Beecher, 2016; Vorvoreanu, Gray, Parsons & Rasche, 2017; Töre Yargar, Süner & Gümüş, 2018) and growing interest in integrating UX education in human-computer interaction education (St-Cyr, MacDonald, Churchill, Preece & Bowser, 2018).

Designing for meaningful user experiences (Hassenzahl et al., 2013; Orth, Thurgood & van den Hoven, 2018) without a significant approach to follow could be vague for designers. When integrating this phenomenon into design education, it becomes even more challenging to assess students’ work. This paper opens the question of how to assess students’ works that aim to design for user experience. For this, we exemplify the approach we followed to assess the user experience design works and take one of the courses we coordinated in the Netherlands. The context for this paper is to design a bachelor’s course for industrial design students at the University of Twente. In this course, several frameworks and topics related to user experience such as human needs (Deci & Ryan, 2011), product experience (Desmet & Hekkert, 2007), product personality (Govers, Hekkert & Schoormans, 2003) as well as the one we developed in 2014 (Bogazpinar, Bakirlioğlu, Kuru & Erbug, 2014) are delivered. The students were asked to explore the possibilities and design products that would fit the usage scenario they developed. Basically, the course aimed to make students identify the role of product design at different levels of people’s experience and formulate design-based documentation of people’s experience. Taking the learning goals of the course as the core, a rubric was designed that identified the expectations from the assignments. In this paper, the approach and the utilization of rubrics in assessing the user experience design assignments in design education are discussed. The contribution of this paper is twofold: (1) it puts forward the challenges of assessment of user experience assignments, and (2) it demonstrates how rubrics, a commonly used way of assessing creative work, could be a way of resolving these challenges. In the end, we suggest that usage of rubrics for assessment of user experience assignments in design education could be one of the ways to assess the vagueness of user experience work.

2 Method

In order to explain our approach, we will first give the structure of the course as well as the assignment we designed. Then, we will explain and discuss the approach we followed to develop rubrics and assess the student works of user experience with the rubrics we developed.

2.1 Structure of the Course

The education system at the University of Twente is unique within the country with its TOM model that is characterized by project-led education. According to this model, each educational year is divided into four thematic modules, and project-led education is supported with other courses within each module. That is, the core of each 10 weeks’ quarter is the project courses, and the other theory-based courses in the module supports the students with the knowledge and experience they could use during each module. These courses end mostly with written exams, but there are also courses that end with smaller design assignments next to the bigger projects. This approach is not only applied to design education but also to all engineering and social sciences undergraduate programs. Students are expected to apply the knowledge they gain from other components of the module in the projects.

Within this system, Design and Meaning course is a third- and final-year bachelor’s course of the Industrial Design Program. It is one of the theory-based design courses of the program which is integrated into the Systems Engineering module. This course stimulates students to work on their own to fully develop a consumer product by taking the relevant theories into account during the design process. The course is expected to prepare the students for their bachelor assignments as well as their future careers, and it compliments other human-centred design-focused courses of the curriculum. Mainly, the course delivers recent models and frameworks of user experience by focusing mostly on conveying meaning through product design.

During the course, students get acquainted with several theories and frameworks that connect design, meaning and user experience. The course focuses on the role of design at various levels of people’s experience with products, and it consists of theory lectures with exercises and an individual design assignment. The theories and frameworks are examined using literature, assignments and practical work. Throughout the course, students iteratively develop means to analyse, appropriate and generate design following the provided user experience frameworks. The course was conducted in 2018 with the participation of 76 students.

---

1 In University of Twente, study programs consist of modules that comprise of several individual courses with different ECTS. In total, each module has 15 ECTS and to graduate from the study program, each student has to get at least 120 ECTS.
In the course, we formulated a fictional story and asked students to design a physical product for facilitating cryptocurrencies as daily payment mediums. This new type of product would transfer several experiences from the currently used products and would generate new experiences as well. Students were free to decide on the qualities and meanings that the product would convey. However, the given two qualities were being portable and being secure. Students were assigned to submit three assignments: Needs Analysis, Extending Product Meaning, and Design for Holistic User Experience.

The course was offered for seven weeks, and the study load of the course was 2.5 ECTS. Students were expected to spend 72 hours within seven weeks. Thirty-two hours of this period of the total time was spent with lectures, workshops and tutorials, and the assignments were designed in a way that students would not spend more than 40 hours. The tutorials, brainstorming sessions and workshops that were listed as the official lecture hours of the course were designed in a way that the materials students produced during these hours would be used as part of the assignments.

### 2.2 Expected Learning Outcomes

While designing or redesigning a course, an important aspect is a constructive alignment. Constructive alignment means that the learning outcomes are in line with the way they are assessed and the teaching method applied during the course (Biggs, 2014). The clearer the learning outcomes are formulated, the easier it is to guarantee constructive alignment. Because if the final result of the course (the learning outcomes) is clear, it is clear what the students should display at the end of the course. One way of formulating clear learning objectives is to use SMART (Specific, Measurable, Achievable, Realistic and Timebound) objectives; this method is originally developed for management goals, but is also advised to be used in educational settings (Bjerke & Renger, 2017).

In design education, this can be quite a tricky part. Design is very often interpreted as something vague and therefore subjective (Vorvoreanu et al., 2017). One person can describe for example soft or modern in a specific way, and a different person might give a different meaning to these terms. Having clear learning outcomes is the first step, and when zooming into the quality assessment itself, it is essential to look at the reliability and the validity of the assessment (Moskal, Leydens & Pavelich, 2002).

**Table 1. Expected learning outcomes of the assignments**

<table>
<thead>
<tr>
<th>Level</th>
<th>Learning goals</th>
<th>Assignment 1</th>
<th>Assignment 2</th>
<th>Assignment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing</td>
<td>Define the role of product design at different levels.</td>
<td>Identifying, criticizing and designing with the product personality.</td>
<td>Identifying, criticizing and designing with the product personality &amp; product experience model.</td>
<td>Highlighting the product qualities of your design and reflecting on how your product responds to human-related qualities discussed during the Backwards Design Workshop.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identify opportunities to influence product experience through design.</td>
<td>Identifying, criticizing and designing with the product personality.</td>
<td>Identifying, criticizing and designing with the product personality &amp; product experience model.</td>
<td>Writing an organized reflection on the product experience.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Design a consumer product by evaluating the models and frameworks of design and meaning.</td>
<td>Recalling the previous knowledge and skills for product development.</td>
<td>Recalling the previous knowledge and skills for product development.</td>
<td>Showing the product development process in creative visual essay format. Relate your design with the models and frameworks that we have discussed throughout the course, but mainly with the Path to Long-Term Usage Model.</td>
</tr>
<tr>
<td>Applying</td>
<td></td>
<td>Demonstrating critical thinking on people’s experience with the crypto-wallet, that was performed throughout the product development process.</td>
<td>Demonstrating critical thinking on people’s experience with the crypto-wallet, that was performed throughout the product development process.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demosntrating the skills of harmonizing the knowledge in an understandable and clear way.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assessment with high reliability means that the outcome of the assessment by the assessor himself is not influenced. If there are multiple assessors, they assess equally and are not biased by possible personal preferences. The reliability is about objectively assessing only what is being assessed. For example, when assessing a sketched concept of design and the assessment is purely meant to be about the concept, the quality of the sketch does not affect the grade. The last aspect is how transparent the assessment is, it is about whether the students know what they are being assessed for. Students should know what is expected of them before heading into the assessment or making the assignments. A possible method to increase all three of these aspects (reliability, validity and transparency) is the use of rubrics.
Jonsson and Svingby (2007) came to the following three conclusions after their literature review about the usage of rubrics:

“(1) the reliable scoring of performance assessments can be enhanced by the use of rubrics, [...] ; (2) rubrics do not facilitate valid judgment of performance assessments per se. However, valid assessment could be facilitated by using a more comprehensive framework of validity when validating the rubric; (3) rubrics seem to have the potential of promoting learning and/or improve instruction (p. 130).”

As stated above, the usage of a rubric does not automatically contribute to the validity and transparency of assessing. The validity and transparency depend on how the rubric is created and communicated towards the students.

In this specific course, it was highly essential to help the students use their knowledge for creating design work. Therefore, rather than other forms of assessment, students were given visual assignments in order to motivate them to develop their creativity skills. While designing the assignments, the learning goals of each assignment were listed and reflected on how each assignment serves for the main learning goals of the course. Table 1 exhibits the learning goals of each assignment and indicates how the learning goals of the course were achieved through each assignment. These learning outcomes are in line with the learning goals of the industrial design program. Hence, we can conclude that these learning outcomes serve for improving the design and intellectual skills of the graduates. To achieve this, the learning goals were discussed with the program director before the structure of the course was set.

2.3 Rubric Development Process

Just as in every type of educational programme, design education has specific goals that students have to achieve. Students are expected to come to a certain level at identified skills and provide high-quality work. The difficulty in design education is that there is not one right answer to the problems (Cross, 2001) and the last thing educators want is that all students produce the same product. Thus, the goals are the same for each student, but the way students show how they have achieved those goals can differ. This can result in friction when formulating specific assessment criteria. To reduce the tension, and to provide clear assessment criteria for the assessment of user experience assignments, we designed rubrics for each assignment of the course. For that, we followed the rules for developing reliable rubrics. It should be noted that the goal of this paper is not to provide a general and fully-reliable rubric that could be utilized for every user experience assignment, but to provide evidence for quantifying design for user experience assignments by developing rubrics.

2.3.1 Design of Rubrics

In education, rubrics are used as scoring guides in assessment with three features: criteria for evaluation, the definition of each criterion and the strategy for scoring (Popham, 1997). Accordingly, the criteria, the definition of each scoring criterion, and the rules of scoring criteria should be clear. Rubrics are also used and perceived to have a neutral or positive impact on students’ creativity (Haugnes & Russell, 2016). The straightforward process for creating rubrics includes listing the criteria (such as learning outcomes), the scores of quality indicators (such as pass-fail) and the definition of each quality indicators (such as good / no indication of knowledge) (Andrade, 1997).

Consequently, together with the definition of the assignments, one rubric was generated for each. Those were designed with two questions in mind: (1) what do we expect the students to learn through the assignments and (2) how should these learning goals be distributed among the requirements of each assignment. By following guidelines for creating rubrics, we first created a draft version of the rubric. Through discussion on the learning goals of the course, we incorporated those guidelines with the learning goals of each assignment. To make the rubrics clear for both the assessors and the students, another lecturer checked the understandability of the rubrics. Several changes were made before the rubrics were used for grading. A rubric example is given in 2.3.2 Reliability of Rubrics.

In an assessment of design works, different assessors could come up with different conclusions. Increasing the quality of assessment and ensuring consistency are the most mentioned benefits of using rubrics in the assessment process, especially if there is more than one rater (Jonsson & Svingby, 2007). This consistency with the rubrics is measured by inter-rater reliability analysis, and the alpha values for inter-rater reliability analysis above 0.70 are regarded as sufficient (Brown, Glasswell & Harland, 2004).

In order to understand whether the rubrics we created were reliable for grading, we conducted interrater reliability analysis as well. After the students submitted assignments, the coordinator of the course randomly picked three of the assignments and did grading separately with an external assessor. We then came together to discuss the consistency of the sub-grades as well as the overall grade of the assignment. The interrater agreement of the initial
round was 0.83, which is very high for the first round (Brown et al., 2004). We did not continue with a second round of grading for interrater reliability, as the first round showed that the rubrics are very reliable. We discussed the points that we did not agree in the first round. The disagreement was very small, and then we agreed that the rubrics could be used for further grading.

Table 2. It should be noted that before this assignment, students were introduced a model that we developed for exploring the experience of certain products (Bogazpinar et al., 2014; Karahanoğlu & Bakırhoğlu, 2017). Therefore, the students were familiar with the terminology used (e.g. human-related qualities) in the rubric made.

### 2.3.2 Reliability of Rubrics

In an assessment of design works, different assessors could come up with different conclusions. Increasing the quality of assessment and ensuring consistency are the most mentioned benefits of using rubrics in the assessment process, especially if there is more than one rater (Jonsson & Svingby, 2007). This consistency with the rubrics is measured by inter-rater reliability analysis, and the alpha values for inter-rater reliability analysis above 0.70 are regarded as sufficient (Brown, Glasswell & Harland, 2004).

In order to understand whether the rubrics we created were reliable for grading, we conducted interrater reliability analysis as well. After the students submitted assignments, the coordinator of the course randomly picked three of the assignments and did grading separately with an external assessor. We then came together to discuss the consistency of the sub-grades as well as the overall grade of the assignment. The interrater agreement of the initial round was 0.83, which is very high for the first round (Brown et al., 2004). We did not continue with a second round of grading for interrater reliability, as the first round showed that the rubrics are very reliable. We discussed the points that we did not agree in the first round. The disagreement was very small, and then we agreed that the rubrics could be used for further grading.

### 2.4 Validity of Rubrics: Evidence from Results of Assignments

Rubrics allow students to identify the critical components of the assignments, which help them to evaluate their performance and progress, and make the marks fair and transparent (Reddy & Andrade, 2010). To achieve this, the language of the rubrics should be clear and understandable for the students as well. This ensures the validity of the designed rubrics (Reddy & Andrade, 2010). The results of the assignments could be used as evidence of validity (Jonsson & Svingby, 2007) in order to understand (1) whether each subpart of the assignment contributed to the learning goals of the course and (2) whether the rubrics were effective in measuring those goals. To check the consistency of each assignment, we analysed the variance of grades of each assignment. To remind, the students worked on the first assignment as a group and proceeded individually in the second and third assignments. Therefore, the range of the distribution of the grades of Assignment 1 was smaller than the other two assignments (Figure 3).
As can be seen, there were three different subgrades in the first assignment: students had to discuss the models and frameworks we covered during the lectures, define a scenario for a potential of a product to be designed and present all these in a good quality poster. In the end, the final grade was calculated by considering the weight of each part, in line with the learning outcomes (See: Table 1). Cronbach alpha\(^2\) for this assignment was calculated was 0.85, which indicated that each part contributed to the measurement of the learning goals of doing this assignment.

The range of the distribution of the grades of Assignment 2 was more extensive than the first assignment. It is also because the number of assignments graded was more than the first one. As can be seen in Figure 4, the range of the grades of “product personality” and “product experience” was greater than “product development” and “visual representation,” while the mean value of each part was around 7.00. Still, the Cronbach alpha calculated for this assignment was 0.90, which is also high, indicating that each subpart of the assignment contributed to the reliability and validity of the assessment.

\(^2\) Cronbach alpha is a statistical indicator of the reliability of measurements, which could be minimum 0.00 and maximum 1.00; the higher the value, the more a test is reliable.
The situation with the third assignment was almost the same as the second assignment. The range of the distribution of the grades of Assignment 3 was broad, and the mean value of each subpart is around 7.00. Still, the Cronbach alpha calculated for this assignment was much higher (0.92), again indicating that each subpart of the assignment contributed to the reliability and validity of the assessment.

These results showed that each assignment was reliable within itself and contributed to measuring the learning goals of the course; besides, the rubrics that we created for measuring the learning outcomes fitted to understanding whether or not the students achieved the goals of the course (Listed in Table 1). These also indicate that the assignment properly fitted to measuring all levels of stated learning outcomes of the course. On the other hand, the reason why the range of grades for Assignments 2 and 3 was high was that some of the students misunderstood the requirements of the assignments. That was one of the issues that require clarification for the assignments rather than the rubrics.

3 Discussions and Conclusion

In this paper, reflecting on the challenges of assessment of design for user experience assignments, we explain the usage of rubrics we developed for a fictional story for experience assignment. In doing this, we focused on the reliability, validity and transparency of the rubrics. Our findings showed that the reliability of the rubrics we created for each user experience assignment was high, indicating that the definition for each grade reduced the subjectivity in the evaluation. That is to say, the grades given to the students is not biased by the lecturer and if another lecturer would give this course in the future, by using the same assessment criteria, together with the same rubrics, the validity and reliability of the grading would be high.

Since the rubrics were highly detailed in rating the assignments, it was easy to find the evidence of student learning. The rubrics also made it clear for the students on how they were graded. Once the grades were announced with given feedback on the assignments, students were given the opportunity to discuss their grades. Only three of the students, whose works were graded less than satisfactory approached the assessors to get extra feedback. Other than those, none of the students wanted to discuss their grades or ask for additional clarification. This incidence on its own is another evidence of the clarity, validity and transparency of the assessment. A challenge the assessors faced was that filling in the rubrics and adding personal feedback was quite time-consuming. Since the rubrics provide a lot of clarity about what is expected of the students, maybe the use of the rubrics could be taken a step further. In the future, we could also experiment with using rubrics as a tool for self-assessment.

To conclude, through developing a clear rubric for this course, we were able to better convey the expectations for each assignment and overall learning outcomes to students and implemented a fair assessment strategy for design outcomes that is clear and objective. The distribution of grading for each assignment presents the reliability of the assessment in this case, and we believe the rubrics can be adapted for other UX-focused assignments and courses with design outcomes elsewhere, according to the different grading systems. However, although the rubrics presented in this paper were useful instruments for students to understand the learning outcomes of each
assignment, the clarity of assignment descriptions is also an important factor for fair assessment of UX-related design work.

References


About the Authors

Armağan Karahanoğlu, PhD is an assistant professor at the Design, Production and Management Department of University of Twente, Enschede, Netherlands, with a special focus on interaction design. Her research interests are design for behaviour change, effects of the products on behaviour, and design for experiences of interactive product/systems.
Charlotte Oude Alink, MSc is educational advisor for the Faculty of Engineering Technology, University of Twente. She holds a BSc in teaching and an MSc in Educational Science and Technology. Her fields of expertise are project-led education, student-driven learning, educational innovation and implementation, and teacher development.

Yekta Bakırloğu, PhD is a postdoctoral researcher in Design Factors, University of Limerick. He holds a BSc, an MSc and a PhD in industrial design from Middle East Technical University. His research interests are open design, design for sustainability, sustainable production and consumption, and design education for sustainability.