Assessment of Theses in Design Education: Conceptualizing a Co-Creative Process for Grading by Automated Criteria Evaluation

Nina Svenningsson, Kristianstad University, Sweden*

<https://orcid.org/0000-0002-2949-9437>

Montathar Faraon, Kristianstad University, Sweden

<https://orcid.org/0000-0002-9740-2609>

Victor Villavicencio, Art-O-Matic AB, Sweden

<https://orcid.org/0000-0003-4611-4129>

ABSTRACT

This article explores and proposes a design concept of a co-creative process that aims to support the assessment and grading of theses in design education through automated criteria evaluation. The research is based on a concept-driven design approach that theoretically anchored and empirically informed the design concept. The research was achieved by grounding the concept in theoretical resources concerning pedagogical principles and assessment, existing tools, and models for examiners in assessment processes and current design practices for assessment in higher education. The main contribution of this article, namely the concept of grading by automated criteria evaluation (GRACE), aims to provide support and structure for examiners and students to collectively advance the design, implementation, and evaluation of the concept through the co-creation and evaluation of criteria in higher education. GRACE could supplement existing assessment practices of theses in design education by focusing on both explicit criteria and the development of students' design thinking and abilities.

KEYWORDS

Assessment, Co-Creation, Criteria, Design, Grading, Higher Education, Learning, Thesis

INTRODUCTION

In the field of higher education, assessments offer essential information in terms of establishing whether institutions and faculty are effective, and whether students are ready for employment (O'Hagan & Wigglesworth, 2015). The way in which such assessments have been performed has not changed noticeably over past decades, despite the fact that research has shown that current grading systems do not work satisfactorily for either students or examiners (Nilson, 2016).

DOI: 10.4018/IJITLHE.294567

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

An undergraduate thesis is the final piece of work undertaken by a student in the last term of a three-year bachelor's degree. The assessment of undergraduate theses is often based on certain criteria that aim to guide examiners and offer students transparency in the assessment process. Due to the complexity and vagueness of the criteria (Golding et al., 2014; Hand & Clewes, 2000), and examiners' interpretations of these (Bloxham et al., 2016; O'Donovan et al., 2001), it is more common for examiners to use a holistic approach when assessing theses (Bloxham et al., 2016). There is often no consensus among examiners on the criteria, meaning that assessments may be made on an arbitrary basis. Different interpretations of criteria can create variation in grades, which makes it challenging to connect assessments to explicit criteria (Bloxham et al., 2016; Sadler, 2009; Webster et al., 2000).

Research into criterion-referenced assessment has established the importance of *explicit criteria* (Ecclestone, 2001; Hornby, 2003; Woolf, 2004), *mutual understanding* between students and examiners (Bloxham et al., 2011; Ecclestone, 2001; O'Donovan, Price, & Rust, 2004.), the *co-creation of criteria* (Hughes & Cappa, 2007; Pathirage et al., 2007; Rust et al., 2005), and the development of criteria as *a continuous process* for examiners and students (Bloxham et al., 2011; Saunders & Davis, 1998).

There is also a lack of research into how digital assessment tools could be designed for higher education to aid examiners in the co-creation of explicit criteria for the assessment of undergraduate theses and the creation of a mutual understanding between students and examiners in the assessment process. In particular, in the area of design education, there is an additional layer of complexity relating to how the level of creativity is assessed in theses. This means that examiners might have to trust to their own experience and carry out holistic thesis assessments, as some aspects are not explicitly articulated in the relevant criteria.

This article aims to provide a theoretical anchoring and underpinning for a design concept for a digital assessment tool, referred to here as grading by automated criteria evaluation (GRACE). The aim of GRACE is to offer a co-creative assessment tool to facilitate the creation of criteria by students and examiners, and to support examiners in reaching a consensus on how these criteria are applied in practice. This research focuses on bachelor's theses written in the design field, and the research process is guided by the following question: *How can GRACE be conceptualized and theoretically anchored to aid examiners in co-creating assessment criteria for undergraduate theses while concurrently supporting students' learning processes?* The concept-driven design research approach put forward by Stolterman and Wiberg (2010) will be used to address this research question. The scope of this article is limited to a focus on examiners' views and challenges when evaluating and grading undergraduate theses.

BACKGROUND

Pedagogical Principles and Assessment in Higher Education

Over time, institutions have adopted different methods for their assessment processes, and the most common of these are based on the perspectives of norm-referenced and criterion-referenced assessment (Burton, 2006; Lok et al., 2016; Prince, 2016). In norm-referenced assessment, students' grades are distributed over a predetermined range, meaning that specific percentages of a group of students will get, for example, As, Bs, Cs, and so forth. This means that students are graded in comparison to each other to determine their rank, and even in the case where no students have performed at an excellent level, some could still receive a top grade (Burton, 2006).

In criterion-referenced assessment, students are graded against predetermined criteria and standards, which means that they will be assessed on the actual work they have produced (Lok et al., 2016; Prince, 2016). The use of criterion-referenced assessment can result in higher reliability, validity, and transparency (Burton, 2006). In general, neither of these two assessment methods is used to the exclusion of the other (Bloxham, Boyd, & Orr, 2011; Lok et al., 2016). When grading theses,

examiners tend to compare them in order to evaluate whether the correct grades have been given (Bloxham et al., 2011), and this approach is more in line with a norm-referenced process.

Lok et al. (2016) present an approach to working with assessment in higher education that uses both norm-based and criterion-based assessment, and in which these two methods complement each other. Their approach is based on a *feedback loop*, where criteria are set based on assumptions about students' abilities to perform (norm-referencing). These criteria form rubrics that examiners can then use throughout the assessment process (criterion-referencing) and consequently arrive at a grade. When the grading process is complete, examiners have a new set of performance standards, and based on these, they can make new decisions about which should be used to create criteria for the next group of students. Students are assessed against explicitly formulated criteria, meaning that transparency, validity, and fairness are maintained and that the recommendations that favor criterion-referenced assessment processes are followed (Quality Assurance Agency for Higher Education, 2006; Shay, 2008).

Approaches to Assessment and Grading of Theses

There are generally two different approaches used by examiners to decide on a grade when assessing theses: *analytic* and *holistic* approaches (Sadler, 2009). In an analytic approach, examiners evaluate students' work against each criterion and decide on a grade by combining all these parts into a whole. In other words, the analytic approach is based on the existence of explicit criteria. When using the holistic approach, examiners decide on a grade based on their overall impression of the performance by the student in this work (Sadler, 2009). The examiners reach a decision based on their opinion of the work, which gives rise to the possibility of different examiners rewarding different grades for the same work.

There has been a steady shift towards the use of analytic approaches, due to a desire to implement a particular structure and to make the assessment process explicit for students (Sadler, 2009). Despite this, it is not unusual for examiners to apply a holistic approach when grading theses (Bloxham et al., 2011; Mullins & Kiley, 2002; Sadler, 2009). Research shows that criteria are often overlooked in the assessment process (Bloxham et al., 2011; Hand & Clewes, 2000; Mullins & Kiley, 2002; Webster et al., 2000). In some cases, the reason is a lack of understanding and similar interpretation of the criteria among examiners (Sadler, 2009; Webster et al., 2000). This could lead to subconscious interpretations and create variations in assessment and individual judgments based on holistic views of students' work (Bloxham et al., 2016).

The process of writing criteria is a complex one, and a skill that not all teachers have (Lok et al., 2016). By and large, examiners have difficulties distinguishing criteria from each other (Sadler, 2009). The criteria used in higher education assessment processes tend to be vaguely defined (Golding et al., 2014; Hand & Clewes, 2000), and examiners often interpret basic terminology (e.g., the meaning of 'analyze') in different ways (O'Donovan et al., 2001). A consistent understanding among examiners of the standards and phrases used in these criteria is essential to ensure that assessments are reliable (Burton, 2006; Rust et al., 2003). One way to achieve this is to encourage examiners to collectively co-create criteria, as this supports the creation of a mutual understanding of criteria and reduces inconsistencies in assessment processes (Hughes & Cappa, 2007; Pathirage et al., 2007; Rust et al., 2005).

Several studies have recommended that criteria should be made *explicit* (Ecclestone, 2001; Hornby, 2003; Woolf, 2004) and that the understanding of criteria and standards needs to be *mutual* between examiners and students (Ecclestone, 2001; O'Donovan et al., 2004). Explicit criteria and the associated transparency that they offer may open the way for a more democratic assessment process, as this means the process is exposed and described in detail (Ecclestone, 2001). The use of explicit criteria also provides an opportunity to share learning outcomes and develop an understanding of these criteria outside of the "'community' of academics" (Ecclestone, 2001, p. 302).

Although the use of criteria has led students to become more aware of how they are graded, or how they are supposed to be graded (Bettany-Saltikov et al., 2009; Hornby, 2003; Sadler, 2009), several studies have shown that *written criteria are not sufficient* for consensus to be reached between examiners (Bloxham et al., 2016; Burton, 2006; Rust et al., 2003; Sadler, 2009; Woolf, 2004). Bloxham et al. (2016) describe the difficulties in “capturing all the nuances of assessor meaning” (p. 477), as examiners subconsciously weigh and value criteria differently. The examiners’ different interpretations of the criteria may create variations in grades and make it difficult for others to connect these differences to explicit criteria.

The complexity of the assessment process in higher education will arguably create variations in grading by examiners (Bloxham et al., 2016). One possible way forward is to achieve a mutual understanding and ownership of criteria amongst examiners and students (Bloxham et al., 2011). In addition, Saunders and Davis (1998) have emphasized the importance of revisiting criteria and ensuring that the development of criteria is a *continuous process*. When students take part in this process, they also gain ownership, leading to a deeper understanding, more consistency, and less sense of unfairness (MacLellan, 2001; Nicol, 2010; Woolf, 2004). Examiners and students need to have a mutual understanding of criteria for an assessment process to be reliable and valid (Bloxham et al., 2011). Due to the complexity of assessing theses in higher education, there is no assurance that examiners will interpret written descriptions of criteria in the same way, regardless of how well articulated they are (Bloxham et al., 2016; Sadler, 2009). Rather than using written remarks, if examiners adopt verbal communication to a greater extent, they can *continuously* build an understanding of the social consensus that guides the assessment and grading processes (Bloxham et al., 2011).

Prior research suggests that the creation of rubrics and other performance descriptions to support students’ understanding of learning goals and criteria can enhance their motivation and academic results (Brookhart & Chen, 2015). Students’ motivation and results improve when they have a better understanding of what they are supposed to achieve (Ecclestone, 2001; Rust et al., 2003). When students are presented with criteria in advance and are supported in the process of understanding them, they create work that is of a higher standard and are more satisfied with the work they produce (Rust et al., 2003).

In summary, the present section has described the concepts of *explicit criteria*, a *mutual understanding* between students and examiners, the importance of *co-creating criteria*, and the development of criteria as a *continuous process* involving examiners and students. These ideas provide a structure and explanatory value for the concept of GRACE. The use of explicit criteria and the creation of a mutual understanding can contribute to the development of clear criteria that are shared between students and examiners through social processes. The co-creation of criteria and the use of a continuous process to further them theoretically anchor and prescribe core functionalities within the concept of GRACE. The process of assessing theses in higher education is a complex one. There is a need for a framework supported by digital assessment tools that can help examiners and students gain a knowledge and understanding of the assessment process and the criteria against which the assessment is being made.

Existing Tools and Models for Assessing Undergraduate Theses

There are several tools and models that are currently used in the process of assessing undergraduate theses. Examples of such tools and models include *grading schemas* (Hornby, 2003; Sadler, 2009), *rubrics* (Neil et al., 1999; O’Donovan et al., 2001; Prins et al., 2016; Sadler, 2009), *grading models* (De Miguel, 2010), and *assessment process models* (Rust et al., 2005). A common goal for these tools is the articulation of precise criteria that can guide examiners and students to understand what is expected at different grading levels.

Hornby (2003) presents a grading scheme based on *grade-related criteria*. This scheme was developed for application throughout an undergraduate program, and includes both *standards* and *levels*, thus creating an opportunity to follow students’ performance throughout their entire education.

The standards are described using explicit criteria that delineate the expectations in terms of the student's performance. The levels relate to expectations of learning outcomes at the different stages of degree and postgraduate programs, and should be clearly stated for both students and examiners. This grading scheme lacks articulated social processes around the creation and understanding of criteria, thus creating the possibility for variation in the interpretations of criteria among examiners, which may jeopardize reliability and consistency in the assessment process (Bloxham et al., 2016; Hornby, 2003; Rust et al., 2005; Rust et al., 2003; Woolf, 2004).

Sadler (2009) describes two grading schemas that are specifically relevant to criterion-based assessment: *rate-weight-sum* and *rubrics*. In the rate-weight-sum model, each criterion is given a weight, and an examiner then assesses students' work on a numerical scale, i.e., by setting a value for each criterion. Rubrics are a "...*cross-tabulation of criteria against so-called 'standards'*" (Sadler, 2009, p. 163). An examiner uses these rubrics as a guide and assigns each criterion with a level or a standard that matches a student's performance. Rubrics intended for theses in higher education can quickly become complex, and it is hard to get a visual overview of them (Neil et al., 1999). Students have also expressed concerns that although rubrics make more criteria understandable, there is room for different interpretations of the criteria. Hence, rubrics can be a helpful guide in the process of assessment, but only as one part of a solution (O'Donovan et al., 2001).

Rust et al. (2005) argue for a *social constructivist assessment process model*, in which active engagement in the criteria is required from both students and examiners in order to gain a deeper understanding of the criteria. Active engagement can be achieved through grading exercises, peer grading, and discussions (O'Donovan et al., 2001; Rust et al., 2005; Woolf, 2004). The model argues for the *participation* of examiners in the *co-creation* of criteria and an *active engagement process* that involves students giving feedback on the examiners (Rust et al., 2005). The responsibilities of the examiners are based on a process that includes (1) designing an assessment process and co-creating explicit criteria, (2) discussing the proposed criteria amongst examiners, (3) training staff on the assessment process, and (4) deciding on grades and moderation of criteria (Rust et al., 2005, p. 236). From the students' perspective, there is a similar process that includes (1) presentation of explicit criteria, (2) active engagement with criteria, (3) completion and submission of thesis work, and (4) active engagement when giving feedback (Rust et al., 2005, p. 236).

While the above tools and models are useful in their own way, there is a need to supplement them with additional aspects such as *social processes*, *active engagement*, and *explicit communication* in order to support the development of mutual understanding among examiners as well as between examiners and students. The concept of GRACE was designed to integrate the abovementioned characteristics to support collaboration and co-creation of assessment criteria for theses through consensus-seeking processes among examiners. The aim of GRACE is to provide the necessary means for examiners and students to discuss the co-created criteria in *workshops* on a yearly basis. The purpose of these workshops is to create transparent discussions in order to reach a mutual understanding of the level at which students are expected to perform to achieve their desired grades (Bettany-Saltikov et al., 2009; Ecclestone, 2001; Hornby, 2003; Sadler, 2009).

The social processes described above allow for the evaluation and revision of assessment criteria over time to adapt to societal transformations and challenges within individual educational disciplines. In addition to these social processes, GRACE can act as a digital assessment tool that allows examiners to assess a thesis based on a set of criteria and provide real-time previews of how the assessment of each criteria affects the final grade.

Design for Assessment in Higher Education

Design has become increasingly important in the field of education. Researchers in the field of education have started to embrace design in their research processes, and this has allowed them to advance and create, for example, prototype solutions for virtual learning environments and designs for curricula or software that offer a more direct impact on education (Edelson, 2002; Reeves et al., 2011).

Edelson (2002) has identified three reasons for the incorporation of design when developing educational theories. Firstly, design as a research practice offers a productive process through which alternate views and inconsistencies can become more visible. In addition, a design process may be research-oriented, which provides a focus for theory development. Secondly, a research process with a focus on generating research products that can be applied in the education domain is closely linked to the design core of education (e.g., the design of educational materials or systems). Finally, design research offers the possibility for innovation, which Edelson (2002) argues is instrumental to solving existing problems or creating opportunities in educational systems.

A growing number of studies have applied design to the development of education in different contexts. For instance, Nilsson et al. (2009) used design thinking to develop mathematical activities for learning in outdoor settings. Cerratto-Pargman et al. (2012) explored different design approaches in an attempt to develop the educational field from the perspective of how information and communication technologies (ICT) impact society. Another contribution to the area of ICT is the framework mLearn4web, which was designed to integrate ICTs into education (Zbick et al., 2015). This tool enables educators to design learning activities and to deploy them on cross-platform mobile applications.

In the context of assessment, Deeley (2018) examined the ways in which different digital technologies could be used for self-assessment. The results showed that digital technology could be beneficial for developing summative feedback assessments with a focus on learning. More specifically, the results showed that the use of digital technology could improve feedback, deliver it faster, encourage more dialogue between students and teachers, and create an individualized relationship between students and teachers (Deeley, 2018).

Although many assessment systems have been designed over the years, most have used a qualitative approach (Valenti et al., 2003). GRACE aims to support qualitative and quantitative assessment methods and to promote new ways of assessing undergraduate theses in higher education. For example, various scales or measurements could be used to evaluate a student's performance at a thesis seminar, and this evaluation could be included in a final assessment.

METHODOLOGY

Concept-Driven Design Research

There are several research approaches that are available in the field of interaction design. Stolterman and Wiberg (2010) describe a few of these: user-centered design, participatory design, contextual design, activity theory, and ethnographically informed system design (p. 97). Many of the existing interaction design methodologies are based on empirical foundations, meaning that their purpose is to explore and analyze users, examine specific situations, and arrive at a design proposal for a particular problem or situation (Stolterman & Wiberg, 2010). These approaches are also being used in design research. However, they lack a focus on theory, which means that they fail "to develop more conceptual and theoretical contributions to the general understanding of the interaction between humans and digital artifacts" (Stolterman & Wiberg, 2010, p. 98).

Stolterman and Wiberg's (2010) idea of concept-driven interaction design research is a complementary methodology that focuses on theoretical advancements. The method is differentiated from other research methodologies in that it has a theoretical point of departure and is directed toward futuristic scenarios rather than existing user situations. In addition, the design concept or artifact produced by the research process is equally as crucial as the theoretical advancements, as it can express other dimensions of a problem or a situation. A concept-driven design research methodology is distinguished by the following properties (Stolterman & Wiberg, 2010, p. 98):

1. The point of departure is conceptual/theoretical rather than empirical.

2. The research furthers conceptual and theoretical explorations through the hands-on design and development of artifacts.
3. The end result (that is, the final design) is optimized in relation to a specific idea, concept, or theory rather than to a specific problem, user, or a particular use context.

Figure 1. The process of the concept-driven design research approach in relation to theory and the use situation, adopted from Stolterman and Wiberg (2010, p. 101)

Due to the complex and changing nature of design and design research, there is a need to explore theoretical advancements with the help of design and through the creation of artifacts (Stolterman & Wiberg, 2010). As previously mentioned, the focus is on futuristic scenarios and the prospects of building on existing theory to support theoretical advancements. Although the research process leads to a concept design, a successful concept is not in itself the primary goal of the approach (and is not necessarily important). Instead, it is “the knowledge production in the form of theoretical developments” (Stolterman & Wiberg, 2010, p. 101) that is the basis for a measurement of success. The intent is not to build a prototype, which describes a solution to a problem, but rather to “address and challenge existing theoretical concepts and frameworks” (Stolterman & Wiberg, 2010, p. 104).

Stolterman and Wiberg (2010) describe seven methodological activities of concept-driven design research: *concept generation*, *concept exploration*, *internal concept critique*, *design of artifacts*, *external design critique*, *concept revisited*, and *concept contextualization*. In the first activity, *concept generation*, existing theoretical work in the field inspires new concepts. These concepts can be developed through different activities and methods, and the goal is to produce something new. In the second activity, *concept exploration*, the aim is to develop a particular concept further. The researcher(s) often work by hand, experimenting with materials to build models and prototypes. During the third activity, *internal concept critique*, the concepts are examined to establish (1) the uniqueness of the chosen core concepts; (2) the extent to which they relate to existing theory; and (3) how well these concepts can be clearly expressed in a concrete design (Stolterman & Wiberg, 2010, p. 110).

The fourth activity, *design of artifacts*, aims to turn the design concept into a concrete artifact. In the fifth activity, *external design critique*, the design of the concept is evaluated by external parties and critiqued to validate both the idea and the theoretical principles, or in other words the design concept as a whole. Following the design critique, the sixth activity, *concept revisited*, gives the researchers a chance to go back and evaluate the concept based on the critiques received in the previous step. Depending on these critiques, different aspects of the concept may require alteration. In the seventh and final activity, *concept contextualization*, the focus is on theory. The proposed

concept is anchored to existing research, and the researchers show how their work contributes to and furthers knowledge in the chosen area.

Examples of studies in which concept-driven design research has been used include those by Faraon (2018), Johansson and Wiberg (2012), Johansson et al. (2015), and Nazzi et al. (2012). Another example is a study by Faraon (2018) that used concept-driven design research to develop the concept of co-creative media to facilitate democratic engagement by citizens. To theoretically anchor the concept of co-creative media, the study implemented concept-driven design research in literature reviews and empirical studies that resulted in four design guidelines. These guidelines aimed to support the development of co-creative media for democratic engagement (Faraon, 2018).

Application of Concept-Driven Design Research

The method of concept-driven design research was chosen here based on its applicability to the research question in this study. The process of this research followed the seven activities set out by Stolterman and Wiberg (2010). For the first two activities, *concept generation* and *concept exploration*, the design concept was articulated in the form of sketches that provided a foundation for the internal design critique. The current body of research provided the understanding necessary to design and develop an assessment process using GRACE. As part of the third activity, *internal design critique*, the design concept was analyzed and theoretically anchored based on contemporary research studies. Five core criteria emerged and were used in the crafting of the initial design concept for GRACE. These were *co-creating criteria*, *explicit criteria*, *a mutual understanding of criteria* between students and examiners, *a continuous social process* for reaching a consensus on criteria, and *a feedback loop* to further the development of criteria.

The co-creation of criteria could provide structure for examiners, as this process encourages them to work together collectively to reach a consensus and a mutual understanding of the criteria (Rust et al., 2005). Mutual understanding can reduce inconsistencies in assessment processes (Hughes & Cappa, 2007; Pathirage et al., 2007; Rust et al., 2005) and enhance their reliability (Burton, 2006; Rust et al., 2003). Following the co-creation of criteria, a workshop or a seminar could be planned with students to make the suggested criteria explicit and to create a better understanding of their expectations of the assessment process (Ecclestone, 2001; Hornby, 2003; Quality Assurance Agency for Higher Education, 2006; Woolf, 2004).

Prior research has shown that students presented with implicit criteria may not understand what they are being assessed on or the requirements for achieving a specific grade (MacLellan, 2001; Rust et al., 2003). However, students who are provided with explicit criteria and are supported to understand them produce high-quality work and feel more satisfied with that work (Rust et al., 2003). Presenting explicit criteria and offering students an opportunity to ask questions about them could lead to a mutual understanding between examiners and students (Bloxham et al., 2011). Inviting students to become a part of a social process that continuously revisits and furthers criteria can enhance the students' feelings of consistency in the assessment process (MacLellan, 2001; Nicol, 2010; Woolf, 2004). The social process requires active engagement and communication in order for both students and examiners to benefit from criteria-based assessment processes (Bloxham et al., 2016; Rust et al., 2005). This feedback loop provides examiners with essential information about the usefulness of current criteria and rubrics, the ways in which they have been understood by students and examiners, whether they need to be clarified, and how they can be adjusted for the next group of students (Lok et al., 2016). The theoretical underpinnings identified in the previous activity were applied in the fourth activity, *design of artifacts*. The outcome of the fourth activity was a first draft of the concept of GRACE, which was presented to external participants for evaluation, see Figure 2.

The principles of GRACE are built upon supporting the social processes that aid examiners and students in co-creating criteria and reaching a common understanding of them, and promote shared expectations of assessment in the grading process. In the first step, examiners work together to co-create and reach consensus on a set of criteria. These criteria are then shared with students

Figure 2. A theoretically anchored design concept for GRACE that combines the characteristics, values, and procedures needed to aid examiners in the co-creation of assessment criteria for undergraduate theses while concurrently supporting students' learning processes.

through workshops, in order to create a common understanding of them and to support the students' understanding of what is expected to reach a desired grade. When the time comes to assess a thesis, an examiner fills in a digital grading form that is based on the agreed-upon criteria. During this assessment, a real-time preview lets the examiner know how the assessment of specific criteria will affect the final grade. When the assessment is finalized, a student receives the given grade together with documentation on how each criterion was assessed.

The fifth activity, *external design critique*, was conducted during a three-hour design workshop with participants at a medium-sized university located in the south of Sweden. There were six

participants, who were faculty members undertaking teaching and research in relation to a three-year design course. The meeting started with a presentation about the concept, in order to describe and articulate the specific characteristics of the digital assessment tool. Following this presentation, the participants were allowed to express their opinions about the concept. They were asked to reflect on the advantages and disadvantages of using such a digital assessment tool, and whether they could see anything that was missing or needed adjustment. The reflection session provided each participant with the opportunity to reflect and articulate their thoughts about the design concept. The session developed into an elaborate discussion that led to new topics and descriptions, often using metaphors. The discussion focused on how criteria could be articulated for the assessment of design and what should be done to foster and encourage design thinking in students. Part of the discussion focused on the difficulties students often have in grasping the process of writing design-oriented academic theses. The sixth and seventh activities, *concept revision* and *contextualization*, are described in the section entitled “Revision and Contextualization of GRACE”, which follows the section in which the results are presented. In total, all seven activities were completed over a period of six months.

Ethical Considerations

We took into account the ethical considerations outlined by the Swedish Research Council (2017) during data collection, i.e., the external design critique activity in the concept-driven approach. The four requirements are *information*, *consent*, *confidentiality*, and *usage*. Firstly, participants were informed about the purpose of the study, and were told that participation was on a voluntary basis and that their responses would be anonymized. Secondly, participants were also informed that they could choose to end their participation at any time during the meeting. Finally, they were informed that the data collected would solely be used in the current study and would not be shared with any third party.

ELABORATING THE CONCEPT OF GRACE

Two themes emerged when the information from the external design critique was compiled: (1) how to assess for design and acknowledge creativity, and (2) how to encourage holistic and subjective assessment without arbitration. The six participants will be represented as P1–P6 to retain their anonymity in the following discussion.

How to Assess for Design and Acknowledge Creativity

The result of a design process is not something that is (or at least should be) predictable: the outcome should be expected to be unexpected. In relation to assessment, this means that the conventional use of criteria in higher education might not be the most appropriate way to assess specific aspects and parts of design theses. Three participants reflected on the importance of evaluating the novelty and the level of innovation in design theses in the following way:

“The unique factor is important in a design process, and we should not accept the world as it is. How do we handle this unknown in assessment?” (P2)

“How should we articulate criteria about the expected unexpected? Is it possible? What might be lost on the way?” (P3)

“How do we encourage students to live in uncertainty about what the result can be and trust the journey? Can we still do this while presenting them with specific criteria and, for example, ask them to hand in suggestions for research questions that are addressed systematically?” (P6)

Design knowledge is not incremental in the same way as in other academic fields. A student may propose a novel and highly creative idea without, for example, referring to or even understanding the

basic guidelines of design theory. For GRACE to consider these aspects, the possibilities of tailoring the assessment process to different fields or needs should be explored further.

Moreover, there were concerns among the participants that the use of fixed criteria might mean that students had difficulty distinguishing their own performance from those of fellow students. There is a risk that students who follow the minimum guidelines and those who produce a thesis with an extraordinary research question, or a creative solution may end up with the same grade. This situation may occur if the criteria measure the fundamentals and the different elements of theses, but not the inspiration, creativity, or excellence in the subject and the ideas behind the outcomes. One participant noted the following regarding this challenge:

“The holistic view feels important, especially in design, where there are often intangible characteristics or thoughts that matter, and things arise/are created that do not match any criteria.” (P1)

Further discussion indicated that students who achieve extraordinary performance should be rewarded. There is a need for criteria or other articulated methods of assessment that will not let extraordinary ideas go unrecognized. One participant provided the following feedback regarding criteria for design:

“Design is about creating something which does not exist; how can we articulate this in a criterion?” (P2)

Another participant proposed a different approach to the assessment of design and suggested the following:

“Assessment needs to be in two parts: one that is more fixed, with criteria for research and scientific content, which applies to all research, and then the design part.” (P3)

The feedback from participants suggests a tension between academic positions that emphasize science and views that focus on design. This may call for a novel view that integrates and combines these two approaches with each other. One possible way forward would be to separate the assessment of design theses into two parts, one of which focuses on criteria and the fundamentals of science/research, and the other on creativity and design thinking.

How to Encourage Holistic and Subjective Assessment without Arbitration

The second theme was an extension of the first. Participants reported that subjective assessments could be gradually standardized between examiners through a continuous shared understanding and interpretation of criteria. The discussion touched upon how examiners could use their expertise in their respective fields to identify ideas and solutions that are unique and have a deeper level of understanding and creativity. One participant emphasized the following:

“There are times when you encounter something truly genius that has never existed before and has never been thought of before. Thus, there is nothing in the applicable criteria that could be used to assess it, but the examiner knows that this is genius. What happens when we cannot use this professional knowledge in the assessment process?” (P4)

Another participant reflected on whether and how criteria could be formulated to facilitate the assessment of tacit knowledge:

“In GRACE, the social processes allow for tacit knowledge to be shared during the design/creation of criteria, but still, examiners can only express what they know—not what they think that they may encounter, in any case not in detail.” (P5)

This feedback implies that examiners could evaluate undergraduate theses according to their expertise and experience gained from years of practice in their specific fields. The instinct of an examiner regarding whether a student has performed something extraordinary should mean something in the grading process; even though examiners might not always be able to articulate what has happened, they “just know” that this is excellent, original work.

Revision and Contextualization of GRACE

Following the external design critique, the design concept of GRACE was revisited and developed further based on the feedback from the participants. Rather than focusing exclusively on criteria, the design concept was redesigned to give examiners support in terms of articulating and designing their assessment process. The formulation and co-creation of criteria is an integral part of this process; however, GRACE also recognizes that different fields may have different requirements and standards for desired outcomes and good work. These standards could be difficult to evaluate in the same way, based on the same types of criteria.

On the one hand, explicit criteria may be suitable for assessing certain parts of a design thesis; on the other, there might be a need to evaluate the novelty and creativity of a design based on different assessment structures that are open to subjective views informed by examiners’ expertise in specific fields. Social processes remain a critical factor in the concept of GRACE. Through social processes, examiners have the possibility of co-creating an assessment process that is suitable for students in specific education and research fields such as design. The design of an assessment process and a grading methodology for thesis work starts with examiners, see the first process in Figure 3.

The process begins with examiners co-creating criteria to reach a consensus about which types of requirements and grading methods might be suitable for assessing theses in their field of education and research. For example, if a thesis course aims to develop creativity or innovative results, then the requirements and grading may focus on the development of new, unknown combinations of ideas; however, if a course focuses on pitching and communicating design ideas, then the requirements and grading could favor visual aesthetics and understandable communication. The outcome is articulated and saved in a document that is available for download by both students and examiners.

When the criteria and requirements for an assessment process have been established, examiners need to convey them to students to promote a shared understanding of expectations regarding a thesis. The key to these social processes is to allow examiners and students to work with the criteria in an active way. A workshop may include various exercises that enable students to highlight parts of sample theses and to discuss them with examiners. The goal is for both students and examiners to go into the thesis work with a broader understanding of expectations and a clearer view of the applied criteria. Upon submission, students upload their thesis to GRACE’s database, and from there, GRACE is expected to aid examiners in the grading process by providing an overview of the requirements. The document containing the criteria is available throughout the whole process, alongside a grade form with the appropriate input choices for each criterion and requirement. As an examiner is grading a thesis, he or she can simultaneously see a preview of the final grade document, which indicates how the final grade is coming together and enables adjustments through the weighting of criteria.

When an examiner has completed the form, the grade becomes final. The grade package delivered to students consists of their submission, a pdf containing the criteria, and the completed grade form, which gives students a transparent and complete understanding of how their grade was assessed and compiled. Following this, students need to be provided with feedback on the assessment process, the applied criteria, and other relevant requirements. This feedback is saved so that examiners can use it when the loop restarts, and it is time to revisit the applied criteria and develop the assessment process

Figure 3. A revised design concept for GRACE that is theoretically anchored and empirically informed by participants in the field of higher education

for the upcoming year. Statistics will be generated from annual thesis submissions and may contribute to developing courses and assessments for forthcoming years. Examiners could use these statistics to understand possible development areas in earlier courses related to academic research writing.

GRACE is a digital tool that has the potential to be shared between different institutions and universities, which would create the possibility of receiving critical feedback. This may also lead to a development process in which GRACE could be implemented horizontally across different institutions and extended by considering the cultural differences in assessment between various research fields. The expectation is to develop a shared understanding of assessment methods for theses, both locally and across different research fields, institutions, and universities.

CONCLUSION

In design education, it is essential to avoid focusing solely on explicit criteria when evaluating the scientific quality of a thesis; it is equally important to consider the development of students' design thinking and their ability to think outside the box. To allow assessment to work well in the design field, there may be benefits from articulating a guide for examiners for assessing specific traits of the design profession, such as creativity, uniqueness, and the novelty of ideas. There is also a challenge in terms of articulating criteria and designing an assessment process that encourages students to think

freely, have the courage to trust the journey that is the design process, and not to jump to conclusions or stick to a specific idea too soon in the research process. GRACE aims to address these concerns and could unify how examiners and students perceive the assessment process for theses in higher education. The findings in this article have drawn attention to the challenges associated with such assessment processes because different academic fields have different needs and specifications in terms of criteria.

Social processes are vital in order to ensure that examiners and students can achieve a mutual understanding on the criteria and assessment process. Co-creation of criteria could be considered a social process that can assist in creating a shared understanding of criteria and reduce inconsistencies in the assessment process. GRACE focuses on the possibility of supporting the assessment of different types and measurement of theses by creating opportunities for examiners to engage in social processes with the aim of collectively articulating criteria and designing their assessment processes. For a group of examiners to get their views across and explain different aspects of the assessment of design, they may need to use metaphors or examples from their own and other related fields. Our findings demonstrated a lack of clear articulation in terms of expressing design assessments and recognizing students' design thinking, abilities, and creativity, which requires further investigation.

The results of this study contribute to the area of design and assessment in higher education through the proposal of a design concept called GRACE, as a potential way of assessing theses that are both design- and research-oriented. As a next step, we could extend the current research to include students' expectations and challenges when writing undergraduate theses. In future work, GRACE could be designed and evaluated in the form of a prototype in order to explore potential challenges associated with its implementation in higher education. Future research may also investigate how an assessment process could be designed to capture both scientific and design aspects of theses while at the same time limiting arbitrary evaluations. In terms of scaling up GRACE, there is a need to investigate how it could be adopted across universities and how this could lead to a consensus about assessing and grading theses at a national level.

REFERENCES

- Bettany-Saltikov, J., Kilinc, S., & Stow, K. (2009). Bones, boys, bombs and booze: An exploratory study of the reliability of marking dissertations across disciplines. *Assessment & Evaluation in Higher Education*, 34(6), 621–639. doi:10.1080/02602930802302196
- Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education practices. *Studies in Higher Education*, 36(6), 655–670. doi:10.1080/03075071003777716
- Bloxham, S., Den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: Exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466–481. doi:10.1080/02602938.2015.1024607
- Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368. doi:10.1080/00131911.2014.929565
- Burton, K. (2006). Designing criterion-referenced assessment. *Journal of Learning Design*, 1(2), 73–82.
- Cerratto-Pargman, T., Järvelä, S., & Milrad, M. (2012). Designing Nordic technology-enhanced learning. *The Internet and Higher Education*, 15(4), 227–230. doi:10.1016/j.iheduc.2012.05.001
- De Miguel, M. (2010). The evaluation of doctoral thesis. A model proposal. RELIEVE. Revista electrónica de investigación y evaluación educativa. *e-Journal of Educational Research, Assessment and Evaluation*, 16(1), 1–18.
- Deeley, S. J. (2018). Using technology to facilitate effective assessment for learning and feedback in higher education. *Assessment & Evaluation in Higher Education*, 43(3), 439–448. doi:10.1080/02602938.2017.1356906
- Ecclestone, K. (2001). 'I know a 2:1 when I see it': Understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education*, 25(3), 301–313. doi:10.1080/03098770126527
- Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences*, 11(1), 105–121. doi:10.1207/S15327809JLS1101_4
- Faraon, M. (2018). Concept-driven design for democracy: Advancing co-creative media to support citizen participation and democratic engagement. *eJournal of eDemocracy and Open Government*, 10(1), 23–49.
- Golding, C., Sharmini, S., & Lazarovitch, A. (2014). What examiners do: What thesis students should know. *Assessment and Evaluation in Higher Education*, 39(5), 562–576.
- Hand, L., & Clewes, D. (2000). Marking the difference: An investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment & Evaluation in Higher Education*, 25(1), 5–21. doi:10.1080/713611416
- Hornby, W. (2003). Assessing using grade-related criteria: A single currency for universities? *Assessment & Evaluation in Higher Education*, 28(4), 435–454. doi:10.1080/0260293032000066254
- Hughes, C., & Cappa, C. (2007). Developing generic criteria and standards for assessment in law: Processes and (by)products. *Assessment & Evaluation in Higher Education*, 32(4), 417–432. doi:10.1080/02602930600898544
- Johansson, D., Lassnanti, J., & Wiberg, M. (2015). Mobile e-services and open data in e-government processes—Transforming citizen involvement. In G. Anderst-Kotsis (Ed.), *iiWAS '15: Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*. New York, NY: ACM.
- Johansson, D., & Wiberg, M. (2012). Conceptually advancing “Application Mobility” towards design: Applying a concept-driven approach to the design of mobile IT for home care service groups. *International Journal of Ambient Computing and Intelligence*, 4(3), 20–32. doi:10.4018/jaci.2012070102
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840–852. doi:10.1080/02602938.2013.875117
- Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: Compatibility and complementarity. *Assessment & Evaluation in Higher Education*, 41(3), 450–465. doi:10.1080/02602938.2015.1022136

- MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307–318. doi:10.1080/02602930120063466
- Mullins, G., & Kiley, M. (2002). “It’s a PhD, not a Nobel Prize”: How experienced examiners assess research theses. *Studies in Higher Education*, 27(4), 369–386. doi:10.1080/0307507022000011507
- Nazzi, E., Bagalkot, N. L., Nagargoje, A., & Sokoler, T. (2012). Concept-driven interaction design research in the domain of attractive aging: The example of Walky. In P. Rodgers (Ed.), *Articulating design thinking* (pp. 227–245). Libri.
- Neil, D. T., Wadley, D. A., & Phinn, S. R. (1999). A generic framework for criterion-referenced assessment of undergraduate essays. *Journal of Geography in Higher Education*, 23(3), 303–325. doi:10.1080/03098269985263
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. doi:10.1080/02602931003786559
- Nilson, L. B. (2016). *Yes, Virginia, there’s a better way to grade*. Retrieved from <https://www.insidehighered.com/views/2016/01/19/new-ways-grade-more-effectively-essay>
- Nilsson, P., Sollervall, H., & Mildrad, M. (2009). Collaborative design of mathematical activities for learning in an outdoor setting. In *CERME 6 Sixth Conference of European Research in Mathematics Education* (pp. 1101–1110). Lyon, France: Institut National De Recherche Pédagogique.
- O’Donovan, B., Price, M., & Rust, C. (2001). The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid). *Innovations in Education and Teaching International*, 38(1), 74–85. doi:10.1080/147032901300002873
- O’Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, 9(3), 325–335. doi:10.1080/1356251042000216642
- O’Hagan, S. R., & Wigglesworth, G. (2015). Who’s marking my essay? The assessment of non-native-speaker and native-speaker undergraduate essays in an Australian higher education context. *Studies in Higher Education*, 40(9), 1729–1747. doi:10.1080/03075079.2014.896890
- Pathirage, C., Haigh, R., Amaratunga, D., & Baldry, D. (2007). Enhancing the quality and consistency of undergraduate dissertation assessment: A case study. *Quality Assurance in Education*, 15(3), 271–286.
- Prince, R. (2016). Predicting success in higher education: The value of criterion and norm-referenced assessments. *Practitioner Research in Higher Education*, 10(1), 22–38.
- Prins, F. J., de Kleijn, R., & van Tartwijk, J. (2016). Students’ use of a rubric for research theses. *Assessment & Evaluation in Higher Education*, 42(1), 128–150. doi:10.1080/02602938.2015.1085954
- Quality Assurance Agency for Higher Education. (2006). Code of practice for the Assurance of Academic Quality and Standards in Higher Education: Section 6: Assessment of students. Mansfield, UK: Linney Direct.
- Reeves, T. C., McKenney, S., & Herrington, J. (2011). Publishing and perishing: The critical importance of educational design research. *Australasian Journal of Educational Technology*, 27(1), 55–65. doi:10.14742/ajet.982
- Rust, C., O’Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231–240. doi:10.1080/02602930500063819
- Rust, C., Price, M., & O’Donovan, B. (2003). Improving students’ learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147–164. doi:10.1080/02602930301671
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175–194. doi:10.1080/0260293042000264262
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179. doi:10.1080/02602930801956059
- Saunders, M., & Davis, S. (1998). The use of assessment criteria to ensure consistency of marking: Some implications for good practice. *Quality Assurance in Education*, 6(3), 162–171. doi:10.1108/09684889810220465

- Shay, S. (2008). Beyond social constructivist perspectives on assessment: The centering of knowledge. *Teaching in Higher Education*, 13(5), 595–605. doi:10.1080/13562510802334970
- Stolterman, E., & Wiberg, M. (2010). Concept-driven interaction design research. *Human-Computer Interaction*, 25(2), 95–118. doi:10.1080/07370020903586696
- Swedish Research Council. (2017). *Good research practice*. Swedish Research Council.
- Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education*, 2, 319–330. doi:10.28945/331
- Webster, F., Pepper, D., & Jenkins, A. (2000). Assessing the undergraduate dissertation. *Assessment & Evaluation in Higher Education*, 25(1), 71–80.
- Woolf, H. (2004). Assessment criteria: Reflections on current practices. *Assessment & Evaluation in Higher Education*, 29(4), 479–493. doi:10.1080/02602930310001689046
- Zbick, J., Nake, I., Jansen, M., & Milrad, M. (2015). A web-based framework to design and deploy mobile learning activities: Evaluating its usability, learnability and acceptance. In D. G. Sampson (Ed.), *IEEE 15th International Conference on Advanced Learning Technologies (ICALT)* (pp. 88–92). Piscataway, NJ: IEEE.

Nina Svenningsson is a user experience (UX) designer at Humblebee in Gothenburg, Sweden. She holds a B.Sc. in Informatics from Kristianstad University, Sweden. Her research interests include conversational design, the influence of design on the trust of online information, and how differences in the way users consume information in digital channels affect their perception of the world and shape their opinions of people with contradicting views.

Montathar Faraon is an interaction design researcher with an interest in co-creation and participatory processes, ICTs for democracy, sociocultural pedagogy, and concept-driven design. He holds a Ph.D. in Information Society from Stockholm University, Sweden and is an Associate Professor in the Department of Design of Kristianstad University, Sweden.

Victor Villavicencio is a graduate of Media Technology at Södertörn University, Sweden. He was the Swedish work package leader in the INTERREG IVA Central Baltic project IleP (Immigrants Inclusion by eParticipation, 2009–2012). He is a researcher, an experienced graphic designer with a demonstrated history of working in the design industry and has a keen insight and knowledge of social media and digital collaboration tools.