Frontiers of Architectural Research (2020) 9, 148–168

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/foar

Research Article

Reinventing the wheel: A tool for design quality evaluation in architecture

Buthayna Eilouti

Architectural Engineering, Prince Sultan University, Riyadh, Saudi Arabia

Received 7 May 2019; received in revised form 16 July 2019; accepted 21 July 2019

KEYWORDS
Design evaluation tool;
Evaluation criteria;
Affective engineering;
Design rubrics;
Architectural assessment;
Design quality indicator;
Quality assurance

Abstract Addressing design evaluation as a critical yet under-researched domain, this study analyzes the design evaluation structure and its variables and criteria that guide its outcomes. Within a scope of architectural design in practice and education, this study develops a new tool for design quality evaluation. The study consists of four main parts. The first is explorative which reviews and analyzes literature in the assessment and evaluation fields. The second is derivative and used to develop a tool for design quality evaluation (DQE) that combines design criteria and detailed evaluation rubrics. The tool which is named "Evaluation Wheel" is intended to help designers, educators and other design-related stakeholders judge design products succinctly, comprehensively, systematically and graphically. The third part is experimental and is used to test the applicability of the proposed wheel. A set of design products are evaluated by a group of designers using two methods; one uses various design criteria, and the other uses the wheel tool introduced in this research. The fourth part is analytical and evaluative which uses deductive, inductive and abductive reasoning techniques to analyze and compare the findings of the third part. The findings are presented and discussed. The results that indicate a reduction of the discrepancies of the scores among the evaluators seem to support the uptake of the proposed wheel in the design evaluation fields.
© 2019 Higher Education Press Limited Company. Production and hosting by Elsevier B.V. on behalf of KeAi. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Representing a continuous interplay between cognition and action (Eilouti, 2018a, 2018b, 2019), intuition and logic, and conceptual and actual, design is one of the most difficult skills to teach and learn. This can be partially explained by its ill-structured nature (e.g. Buchanan, 1992; Eilouti, 2009; Farrell and Hooker, 2013), the multiple intangible ingredients of its construct (e.g. Nikander and Liikkanen, 2014; Volker et al., 2008), and the various implicit layers and tacit heuristics underlying its knowledge resources (e.g. Clancey, 1985; Goldstein et al., 2001; Fu et al., 2015).

E-mail address: beilouti@psu.edu.sa.
Peer review under responsibility of Southeast University.

https://doi.org/10.1016/j.foar.2019.07.003
2095-2635/© 2019 Higher Education Press Limited Company. Production and hosting by Elsevier B.V. on behalf of KeAi. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Designers, and especially novice practitioners, frequently find themselves confront situations where they have to decode design criteria in light of a feedback from a stakeholder, or based on a new scenario that emerges during the designing process. In this context, designers are expected to significantly benefit from making such criteria explicit, measurable, externalized and systematized.

Numerous studies have investigated the creative nature of design within a problem-solving approach (e.g. Cross, 2006; Eilouti, 2007). Within such an approach, Dorst and Cross (2001) relate the cognitive features of experienced designers' thinking processes to problem finding, framing and solving. As a result, they propose a model of creative design that hosts co-evolution of problem and solution spaces. Addressing the same issue, the problem-solving process has been studied in terms of the characteristics that distinguish the design activities of expert and novice designers (Ball et al., 2004). Such a comparison between novice and expert designers' activities does not only reveal the differences in problem-solving techniques but can also provide insights on the evaluation methods of both sides. Furthermore, the evaluation of whether the problem-solving in a given design is successful or not, and to what extent it is successful is necessary for all stakeholders involved in a design.

Evaluation in design education is vital for both sides of teaching: the knowledge sender and receiver. For students, it is important to be informed about their learning levels, strengths, weaknesses, opportunities of improvement, gaps in their knowledge that need additional efforts, and skills that need supplementary sharpening (Hickman, 2007). On the other side, educators also need to judge the effectiveness of their teaching strategies and the progress of their students' learning (Lowe, 1972; Rayment, 2007). Moreover, to improve the effectiveness of evaluation, Thomson (2007) suggests that involving students in the shared evaluation experience is expected to enhance the operative knowledge transformation of assessment standards into quality products. Supporting this viewpoint, Thompson et al. (2017), and López—Pastor and Sicilia—Camacho (2017) found that student involvement in the assessment and evaluation processes is crucial to improving their learning skills. In addition, such an involvement provides a valuable feedback resource to inform the teaching enhancement and development methods (Huxham et al., 2017).

To help facilitate the process of design quality evaluation and to inform the problem-solving activities in design processing, this research explores the major evaluation characteristics and aspects to derive a systemic model of evaluation and to test the impact of the evaluation structure on the final decision about the quality of a given design product. The scope is architectural design in practice and higher education, where various building designs are evaluated using two different methods. The study aims at making evaluation criteria more explicit. It also aims at supplementing the evaluation criteria with a set of more detailed rubrics to help guide evaluators in their assessment and valuation processes, and to help direct designers to solve their design problems and to derive more satisfactory design products.

An externalized and explicit system of design quality evaluation can contribute to the areas of design knowledge, education and practice. Regarding the first area, an externalization of the tacit knowledge of design evaluation criteria has a significant impact on informed knowledge development and application (Lindström, 2006). In education, the development of a new tool that is based on the enhancement of a criteria-referenced assessment system by detailed rubrics can help reduce the differences in the evaluation results of various assessors of the same design (Lindström, 2006). In practice, an explicit system of evaluation can help reduce the gaps between various viewpoints of the stakeholders involved in a design quality judgement. It can also help them in the decision-making process to select stronger candidate designs out of the set of proposals presented to them.

2. Research design

The research consists of four main parts. As it helps establish a knowledgebase of this research, the first part of this study is explorative where it scans the related literature and reviews the main components of the assessment and evaluation acts. The main orientation of the second part is generative. It introduces a novel model for the systematic evaluation of architectural design. The model is presented as a tool that has two versions, one is process-oriented and the other is product-driven. While both versions function as criterion-referenced evaluation tools, the first targets academic applications and the second fits practice purposes. The third part is experimental where it tests the applicability of the model to explore its impacts on evaluation results. The fourth is analytical and evaluative, and uses qualitative and statistical analyses to find out the results of the applicability test and its implications. Hence, the research applies analytical, derivative and evaluative approaches. The first is mainly theoretical and is used in the first part. It is, in turn, applied in conjunction with the professional and academic experience of the author as a point of departure for the derivative approach. This latter is used to develop the model for design evaluation, that is, the tool for design quality evaluation (DQE). The evaluative approach is applied to test the applicability of the model and to reveal the factors that influence the final evaluation outcomes. The evaluative approach is applied in conjunction with a statistical analysis to find out the results of the experimental approach that is used to test the derived model. As such, this research contributes to what Armstrong (1994) calls "authentic assessment" which aims to transform the "factual knowledge" into "procedural knowledge". This latter is seriously needed in the academic field of engineering and architectural design to make the learning guidelines and evaluation bases more explicit and structured. Methodologically, deductive reasoning is used to conclude results from the general observations of the experiment, whereas abductive and inductive reasoning are used to propose criteria for evaluation and to formulate a graphical user-friendly tool for design quality evaluation. Data analysis varies from descriptive qualitative to quantitative measures of design evaluation criteria.

The ontological assumption of the research is that knowledge is expressed through making and is developed through self-evaluation. The epistemological assumption is based on logical positivist where we know what we can test, assess and observe. The methodological assumption is that the structure and method of evaluation influence the results and outcomes of evaluation.

The major contribution of this research is its introduction of an original tool that helps externalize the evaluation criteria that are typically implicit and dependent on evaluator's experience. Such a tool enhances the functionality of evaluation as a part of knowledge development and learning process of designing. The tool is introduced in a visual user-friendly format that lists criteria and sub-criteria and assigns a scale to each evaluation aspect. It also combines qualitative criteria with quantitative scales.

The contribution to the externalization of the tacit evaluation criteria and scales coupled with the recognition that product evaluation is a component of knowledge transfer and development, and a part of the learning process of design and designing, represent the main significance of this study.

3. Evaluation in literature

For the explorative part of this study, a survey of literature in the design evaluation areas reveals multiple aspects of this domain. These include definitions and significance of the evaluation act, and issues associated with the evaluation activity such as its purposes, types, forms, levels, techniques and influential factors.

3.1. Definitions

To define evaluation, it is essential to distinguish between the *evaluation* and *assessment* terms. Assessment-in the education domain-is the systematic process of documenting, testing and using empirical data on the knowledge, competencies, attitudes and beliefs that is used to improve student learning and develop educational programs (Allen, 2004). According to Suskie (2004), assessment helps:

- Establish measurable and clear outcomes for learning.
- Provide sufficient learning opportunities to achieve learning outcomes.
- Implement a systematic method of gathering, analyzing and interpreting evidence to determine the level to which student learning matches educational expectations.
- Objectively understand the state or condition of a setting by observation and measurement.
- Use the collected data to inform future improvements in student learning.

Assessment of teaching means testing and measuring its effectiveness. In this regard, it is possible to classify assessment into three categories, that is, the formative, the summative and the operative types. Sometimes, these are referred to as assessment for, of, or as learning tools. The *formative* assessment -or assessment for learning-is a process of measurement for the purpose of improving the learning effectivity (e.g. Fisher et al., 2011). The *operative* assessment -or assessment as learning-is a process through which students learn new sides about a subject. While formative and operative assessments seem similar, a key difference between the two is the feedback element. In the formative, the educator provides a feedback about the learning process, whereas in the operative assessment the students learn through the process by self-reflection and not necessarily by the intervention of a mentor. The *summative* assessment -or assessment of learning-is what we normally call "evaluation."[1] A common goal of the *summative* type of assessment is to measure the mastery of the expected learning standards. Summative assessment may occur a few times over a course or an academic unit to provide conclusions, whereas the formative and the operative assessments may occur several times during that course to enhance its process.

Evaluation is a value-laden judgment of a program's outcomes, quality, functionality, effectiveness and/or worth. It is a structured interpretation and assignment of meaning to predicted or actual impacts of proposals or results, and a systematic judgement of a subject's merit, quality, value and significance using a set of prescribed criteria, standards and scales. It adds a *value* to the act of assessment. This is usually related to the aim, objectives and/or results of the assessment action. The primary purpose of evaluation, in addition to gaining insights into prior or existing solutions, is to enable reflection and help in the identification of future change and development (Reeve and Peerbhoy, 2007; Scriven, 1967). Moreover, evaluation can be defined as the process of observing and measuring a setting for the purpose of judging its quality and determining its "value," either by comparison to similar settings, or to a set of preset standards. Evaluation of teaching means establishing a measured judgment on its effectiveness as a part of an administrative process.[2] Evaluation is also a rigorous and meticulous application of systematic methods to assess the design, implementation, improvement, and/or outcomes of a program (Ross et al., 2004). It is a resource-intensive process, frequently requiring resources, such as evaluation expertise, labor, time, and budget (Ross et al., 2004).

While both evaluation and assessment include evidence-driven judgmental actions that require data and measures, they differ in many aspects. These differences are listed in Table 1.[3]

3.2. Evaluation significance

Cowan (2000) proposes that asking design students and educators to express in detail the criteria according to which they think that projects are judged, and then comparing the results can be informative to design progress. In agreement with this viewpoint, Armstrong (1994) asserts that what he calls "authentic assessment" is an act that helps shifting the focus of evaluation from the factual knowledge to the ability

[1] Institute for Teaching, Learning & Academic Leadership, University at Albany, www.itlal.org.

[2] Institute for Teaching, Learning & Academic Leadership, University at Albany, www.itlal.org.

[3] Based on an editing of: (www.onlineassessmenttool.com).

Table 1 Comparison between the assessment and evaluation attributes.

	Assessment	Evaluation
Orientation	Process-oriented	Product-oriented
Outcome	Provides feedback	Shows weaknesses and strengths
Value assignment	No value assigned	Assigns values
Reference	Individualized	Compared to prescribed standards
Quality incorporation	Improves quality	Judges quality
Duration	Ongoing	Conclusive and provides closure

to use such knowledge and its associative skills and processes for solving open-ended problems during meaningful tasks. In other words, according to Armstrong (1994) "authentic assessment" and the evaluation associated with it helps transform the "factual knowledge" into "procedural knowledge" and helps bridge cognition to action in design processing (Eilouti, 2018c, 2018d).

Addressing the significant impact of evaluation on knowledge development and on innovative productivity, Lindström (2006) suggests that *criterion-referenced* assessment can help articulate some aspects of the tacit knowledge and emphasize some processual dimensions of the creative work. To achieve this, Lindström (2006) asserts that a criterion-referenced evaluation should be applied in the form of a detailed rubric to be successful. Lindström's research on the validity and reliability of criterion-referenced evaluation shows that given the same rubrics of criteria, when the evaluation of an instructor of an art course and an outside assessor of a similar course are compared, the difference between them is reduced to two steps or less. This helps increase the objectivity and reliability of evaluation. In addition, to increase the effectiveness of evaluation, Thomson (2007) proposes that involving students in the *shared evaluation* experience is expected to improve the operative knowledge transfer of assessment processes and standards into quality products and results.

3.3. Purpose of evaluation

The main purpose of a program evaluation is to determine the quality of that program by formulating a judgment of its performance (Hurteau et al., 2009). In education, evaluation is a measured judgement that helps educators gauge the quality of teaching and learning. As such, unlike assessment, evaluation is a product-oriented rather than a process-centered act. In addition to its quality measurement contribution, Bramley and Newby (1984) identify five main purposes of evaluation:

1. Feedback. Evaluation helps link learning outcomes to objectives and provides a form of reflection for quality improvement.
2. Control. Evaluation links organizational training to pedagogical activities and assists in quality assurance and cost effectiveness considerations.
3. Research. Evaluation provides a knowledgebase that helps determine the relationships between learning, training, and practice.
4. Intervention. The results of an evaluation can be employed to influence its settings and context.
5. Power games. Evaluation results can be used to manipulate organizational politics. This is not necessarily a desired purpose.

3.4. Types of evaluation

In the discussion of evaluation types, it is helpful to start with the types of assessment. These include three major types[4]:

1. A *normative assessment*, or a *norm-referenced assessment*, compares an individual's performance to that of *his or her peers* in a given group. It determines whether an individual achieves at a level below, above or equal to the average performance in a given task for a given group (Glaser, 1963).
2. In a *non-normative assessment* or a *criterion-referenced assessment*, an individual's performance in a given task is judged based on her/his compliance to a given set of standards/criteria that are prescribed according to some expected or desired outcomes.
3. In an *ipsative assessment* system, an individual's performance is compared only to her/his own previous performances.

Notably, it is possible to classify normative and ipsative assessments as relative measures, whereas criterion-referenced assessment as an absolute form of judgement. All of these types have correspondence in evaluation. Hence, it is possible to classify evaluation into normative, non-normative and ipsative types.

In addition to these three types, it is possible to identify two other categories, that is, the formative and the summative assessments that were described in section 1.1 of the evaluation definition. Formative assessment provides an overview of learners' level in the beginning of or during an instruction session. As such, it provides an opportunity to improve the instruction effectiveness. In contrast, a summative assessment is conducted at the end of an instruction session and focuses on the outcomes and the final results of the instruction system.

Similar to the assessment types, Scriven (1967) first suggested a distinction between formative evaluation and summative evaluation. Saettler (1990) defines these two types of evaluations as: formative evaluation is used to refine goals and devise strategies for achieving these goals, whereas a summative evaluation is undertaken to test the validity of a theory or a program or to determine the impact

[4] Institute for Teaching, Learning & Academic Leadership, University at Albany, www.itlal.org.

of an educational practice so that future efforts may be modified or improved. Moreover, a formative evaluation is a method for judging the worth of a program while its activities are still in progress and still forming. This type of evaluation is considered internal and focuses on the process. In contrast, a summative evaluation is a method of judging the worth of a program at the end (summation) of the program activities. This type is external and focuses on the outcomes.[5]

3.5. Forms of evaluation

Jury format is one of the most common rituals of design project evaluation, especially in the architectural design education discipline. It represents the primary interface between critics and learners (Murphy et al., 2012; Webster, 2006). In this format, both evaluation and education are carried out simultaneously in the most well-established performative stage of design education (Webster, 2006). In addition to the jury format, it is possible to identify four other forms of evaluation. These include the one-to-one critique, the peer-evaluation, the on-line evaluation and the anonymous review. In the individualized critiques, an instructor provides feedback and evaluation to each student based on his/her performance. In the peer-evaluation, students evaluate each other's works which provides an effective summative as well as formative feedback and reflections (Gielen and De Wever, 2015; Nicol et al., 2014; Shute, 2008; Snowball and Mostert, 2013). The on-line evaluation is another form of evaluation where the evaluators and the evaluated do not meet face-to-face, but can share products and presentations in the cyberspace in a synchronized or asynchronized manners (Holmes, 2015; Marden et al., 2013; Palmer and Devitt, 2014). The anonymous review is typically used in competitions and conducted to select candidate designs based on prescribed criteria without the presence of the designers (Dijks et al., 2018).

3.6. Techniques of evaluation

One of the evaluation techniques in design research is Saaty's prioritization of the design criteria. This technique encompasses two processes. The focus of the first is on the elicitation of verbally stated attributes. The main task of the second is the numerical scaling of the attributes formulated in the first. Prioritization is a normative ranking technique that is used more frequently in the evaluation of intangible attributes as a means of assigning weights for a set of non-numerical criteria according to their subjective significance (Saaty, 1977; Saaty and Erdener, 1979).

3.7. Levels of evaluation

According to Donald Kirkpatrick (1975), it is possible to identify four levels -or steps-of evaluation. These were later collectively defined as a model that was first published in a series of articles in 1959 in the *Journal of American Society of Training Directors*. Kirkpatrick called these levels as "steps" in a technique for conducting evaluation (Craig, 1996).

The steps/levels of evaluation are:

- Step 1: **Reaction** — This level reveals the feedback and the extent to which the learners like the learning process.
- Step 2: **Learning** - This level highlights the extent to which the learners gain knowledge and learn skills.
- Step 3: **Behavior** - This level emphasizes the capability to perform the newly learned skills and the changes in job performance as a result of the learning process.
- Step 4: **Results** - This level focuses on the tangible results of the learning process including the improved quality, reduced cost, and increased productivity and efficiency.

As the Kirkpatrick's model/technique places the two most important items—results, and behavior— last, it was later flipped upside down to produce a more efficient model (Chyung, 2008; Gilbert, 1998; Markus and Ruvolo, 1990). Thus, the levels of the modified model of evaluation became:

1. **Result** - This level focuses on the impact/outcome/result that can be used to improve the educational system.
2. **Performance** - This level emphasizes the performance of educators and learners in order to create the desired outcomes and results.
3. **Learning** — The concern of this level is the knowledge, skills, and resources that a learner needs in order to perform well.
4. **Motivation** - This level focuses on what learners need to perceive to gain the desired performance.

3.8. Influential factors

Addressing the factors that influence evaluation, numerous studies discuss the various forces that impact the design process assessment methods and the final product evaluation results. Some emphasize the social and qualitative factors in design processing (e.g. Tromp and Hekkert, 2016). The impact of evaluation repetition on the final judgement result was the concern of other studies, such as Coughlan and Mashman (1999) in which they found that the perception of a design concept appears to change with repeated exposures. They concluded that design evaluation protocols which rely on a one-time evaluation may provide misleading information to designers and decision-makers about consumer's satisfaction and enthusiasm for a given design product.

Addressing the varying assessment results of the various stakeholders, Georgiev et al. (2010) found out that users' and designers' viewpoints do not integrate well. To overcome this problem, they proposed a methodology of design factors analysis that is based on meanings as perceived by end users. This methodology introduces a user-derived evaluation of a structure of meaning elements and relations. The analysis of this methodology

[5] (ADDIE Model, www.nwlink.com).

shows the effect of the semantic factors on the resultant evaluation of design. As such, the end users' input and the impact of design on users should be considered as components of the evaluation criteria. In addition, Georgiev et al. (2010) also indicate a difference of design priority between users and evaluators. Their study shows that while users of the case study building were mainly concerned with its functionality, the decision makers emphasized the emotional and intuitive aspects of its design in their selection judgement. The experts seemed more capable than the users to envision the designs in terms of their potential qualities and predicted problems. For experts, the functionality was a basic requirement while the most satisfying designs were the ones that caused excitement, inspiration, pride and love. In this study, the most functional designs could not compete with the more challenging and daring ones. These emotional considerations indicate that the final decision making transcends the constraints of the strictly quantifiable and measurable criteria for justification to the less direct qualitative and normative dimensions.

3.9. Associated issues

In addition to its association with the afore-mentioned problem-solving competencies, evaluation process is strongly associated with decision-making, quality control and creativity issues. In regards to the decision-making skill, Lera (1981) proposes a framework of decision-making and value theory as a result of several empirical and theoretical studies of design judgment. In addition, an extensive body of research on decision-making in design exists (e.g. Ball and Ormerod, 1995; Ball et al., 2004; Girod et al., 2000; Ullman et al., 1996). These studies have brought forward some instances of normative and non-normative aspects of design decisions (e.g. Dorst and Cross, 2001; Guindon, 1990; Jansson and Smith, 1991).

Design judgement aims to assure quality in products. Although it can be defined in various ways, quality is a subjective matter that is based on a set of perceived priorities (Choy and Burke, 2006). The term *quality* is used in various contexts, but mostly in connection with the evaluation of a product or a process (Hubka, 1992). Because quality of architectural design products may be judged from different perspectives, it is difficult to find systematic approaches for its assessment (Van der Voordt and Van Wegen, 2005). Design quality perceptions are analyzed by Volker et al. (2008) for their underlying aspects. These perceptions were compared based on their quality receivers. Thus, the judgments of lay and expert designers were analyzed and compared. The comparison results revealed that the systematic considerations of stakeholders' input and expert evaluation do not preclude a holistic judgment. Furthermore, a variance between the rational and emotional dimensions of architectural design quality assessment was observed. Although the decision makers used a broad range of criteria for design evaluation, the final decision seemed to be based on some emotional perceptions of architectural quality. These perceptions were particularly related to the influential factors that evoked emotional responses and impacted the final decisions.

A multitude of studies focus on a major element of design quality which is creativity. For example, in their explorative creativity assessment indicators study, Demirkan and Afacan (2012) found that three main factors indicate creativity. These consist of the *novelty of design shape*, the elaboration characteristics that are associated with the *geometry* and *figure/ground relations*, and the *compositional factors* including rhythm, repetition, harmony, unity, order and occurrence of design elements.

As a graphic summative overview, the main components of this literature review are illustrated in Fig. 1.

4. Design quality evaluation (DQE) wheel

In this section, a tool of design quality evaluation is proposed, developed, introduced and discussed. The main goals underlying its development are:

- To help make the evaluation act more explicit, systemic, systematic and user-friendly.
- To achieve Armstrong (1994) proposal of "authentic assessment" that helps transform the "factual knowledge" into a "procedural knowledge."
- To help reduce the differences among evaluators' scores as in Lindström (2006) experiment.

A few tools of evaluation have been introduced in the design evaluation literature. One of these is Gann et al. (2003). Although it is an effective tool, it does not cover some important aspects of architectural design, especially in the design education domain. Examples of these missing criteria are the design communication and presentation skills, the design processing and development aspects, and the semantic and concept-related components that typically underlie a design product and implicitly shape its outcomes.

A model for the *summative evaluation* of architectural design products is developed based on a combination of literature review and heuristics derived from the previous academic and professional experience of the author. The tool is developed using iterative cycles, where it started with the main conventional design evaluation criteria such as form, function, contextual fit and structure. Then, with the multiple applications in design studio evaluation and jury judgements, other criteria, detailed sub-criteria and quality indicators were added. The prototypical tool was used for the evaluation of multiple design presentation phases, and it was continuously refined until the final versions introduced here are formulated. The attributes of the proposed model are illustrated in Fig. 2. In this figure, the link between the proposed tool and the previous literature review is illustrated. Although the model is defined as summative, it can be used during the designing process as a formative and ipsative tool. As such, the DQE model can help learners as a self-assessment tool for/of/as learning by design students.

The DQE wheel consists of five blades, each of which represents a major criterion for design quality assurance. The criteria include function, form, context, performance

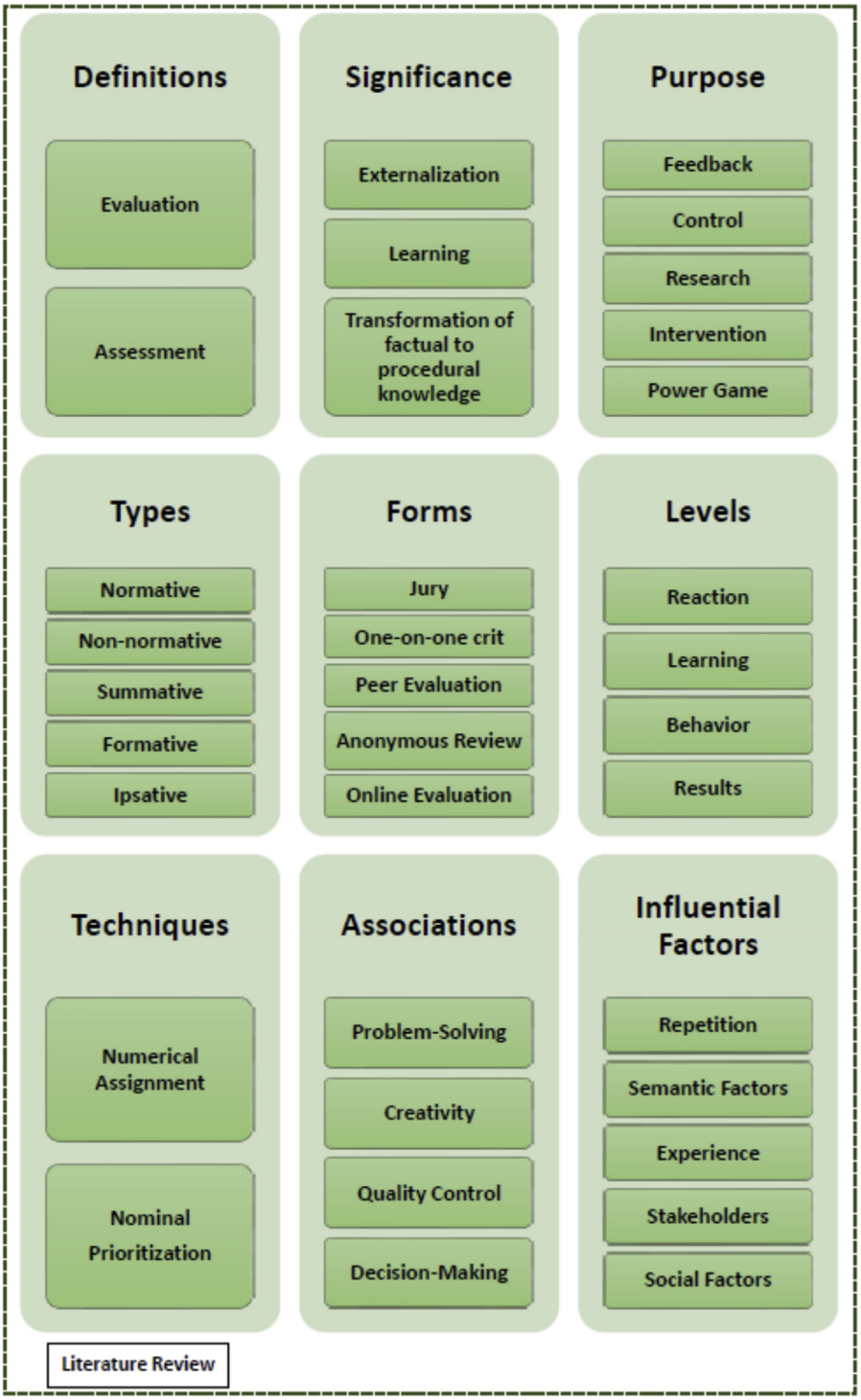


Fig. 1 Design evaluation issues in literature.

and concept. Each criterion, in turn, consists of two main indicators (Fig. 3). The criteria include:

1. *Function* represents the extent to which a building achieves its utilitarian purpose. Its indicators are the space-related configuration quality, and the accessibility solutions. The first indicator is concerned with the quality of each space in terms of its measurements, anthropometrics, ergonomics, orientation, contextual fit and alignment of its geometry to its purpose. In addition to the intra-space attributes that are concerned with characteristics of each individual space, this indicator is also concerned with the inter-spatial aspects such as the spatial layout, topology, organization, proximity relations and the zoning, enclosure and clustering solutions. The second indicator is concerned with the ingress, egress, circulation and barrier-free solutions.

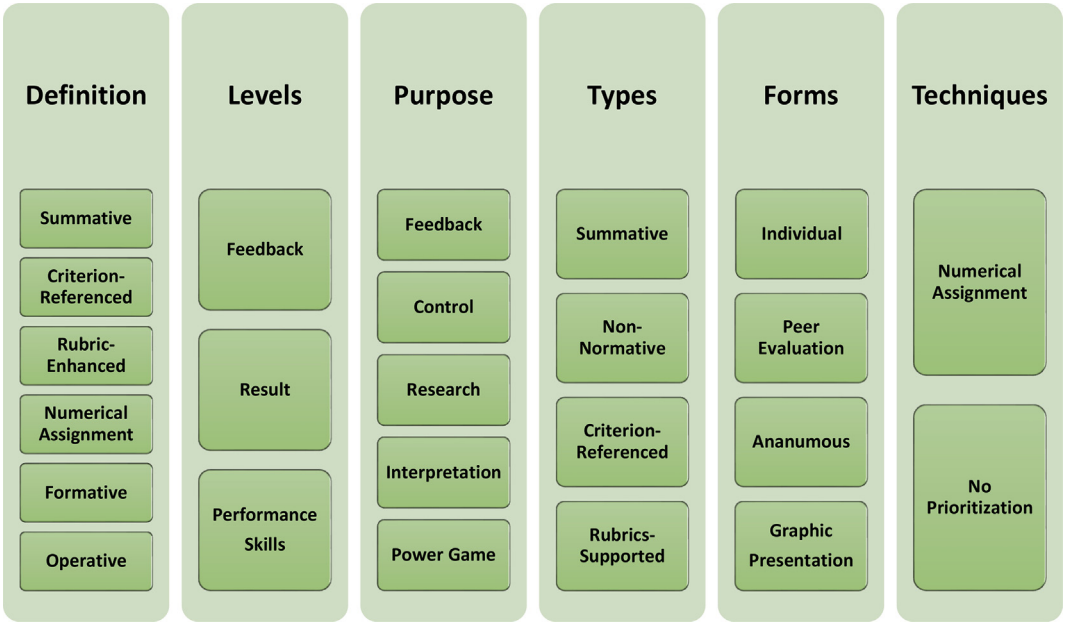


Fig. 2 Attributes of design quality evaluation wheel.

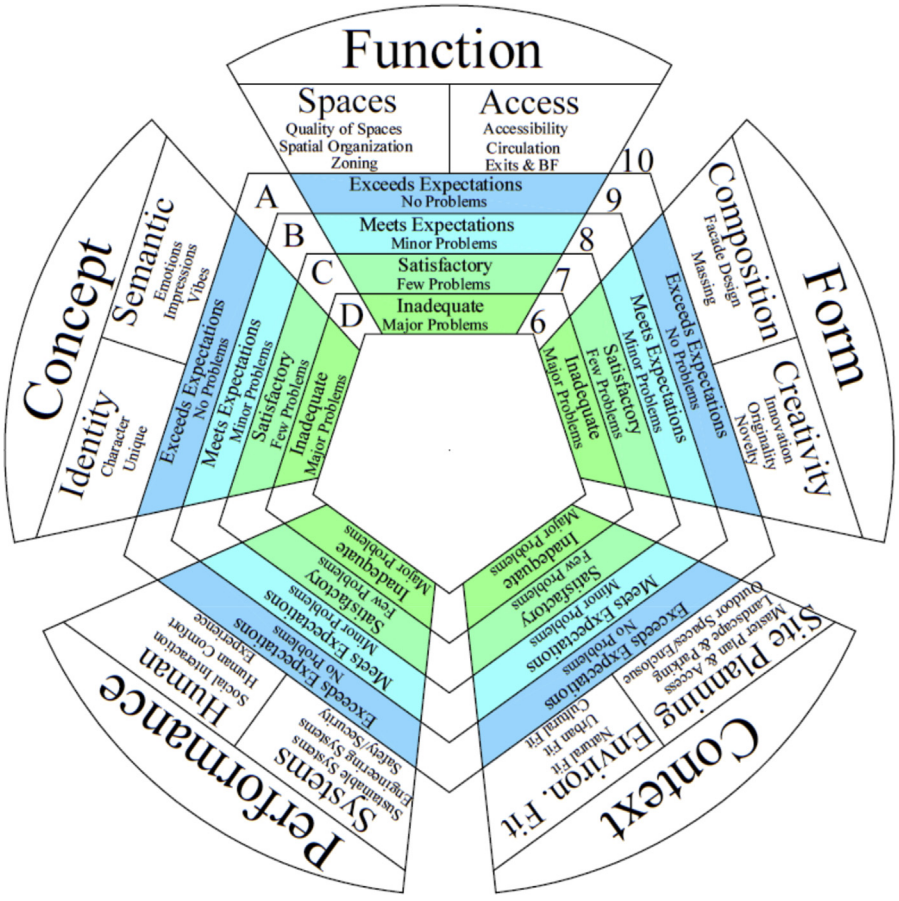


Fig. 3 Design quality evaluation wheel- Product-oriented version.

2. *Form* is represented by the incorporation of the principles of composition and the creativity of the generated designs. The former is usually manifested in the façade design as reflected on each exterior plane of the building, and by its massing and modeling configurations. Both the two-dimensional and the three-dimensional compositions are expected to exhibit attractive articulations to present a successful design. The latter indicator of creativity represents the innovation, originality and novelty aspects of a given design.
3. *Context* is concerned with the quality of site planning and the environmental fitting of the design at hand within its direct and indirect environments. The former indicator includes solutions for the master planning problems, its accessibility schemes, its landscape design and parking allocation, and its outdoor spaces and their spatial enclosure and place-making contributions. The latter indicator evaluates the urban and natural fitting of a building within its surroundings, as well as its cultural fitting within the community it serves.
4. *Performance* of a building can be assessed using two indicators. These are the quality of building systems and the satisfaction of human needs. The former includes the *passive* natural, the *active* engineering systems, and the security, safety, intelligent and management control systems. The building structure is a major part of this criterion. The latter indicator is concerned with the human factors, social interaction, and quality experience a user goes through when using the building inside and outside. This criterion highlights the interior design, building envelop, and outdoor quality aspects of a building. This criterion assesses how well the function is achieved and goes beyond the mere utility of building's function by examining the quality, flow and effectiveness of its tangible and intangible systems. It transcends the functional assignment of spaces and addresses the comfort aspects as well as the spirit of a place that attracts users to stay in or come again to a building, and consequently fulfills the social sustainability requirements of that building.
5. *Concept* represents the *soft* and intangible layer of a design and includes the semantic issues and the uniqueness of a given design. The semantics of a building include the emotions, semiotics, impressions, messages, symbols, philosophy and vibes associated with its design configurations as perceived by users and viewers. The uniqueness is expressed by the identity and character of a building that distinguish it from the remaining designs and make it stand out in a group of buildings in a given context.

The evaluative section of the DQE wheel consists of four concentric pentagons. The outmost one of them represents the highest level (the A-score) of meeting the requirements, criteria and standards of each design criterion and indicator. The concentric pentagon set moves inward from the higher to the lower level that gradually lacks quality and exhibits problems in the design. Consequently, the inmost pentagon represents the D-score which signifies multiple problems in meeting the design requirements and criteria.

As mentioned earlier, Lindström (2006) suggests that a criterion-referenced assessment should be applied as a detailed rubric to help articulate some aspects of the tacit knowledge and to emphasize some processual dimensions of the creative work. Clarity of the rubrics affects the usefulness of evaluation (Kite and Phongsavan, 2017). To elaborate the assessment of the various levels of the wheel, a table of evaluation rubrics is added to supplement the DQE tool, explain its evaluation expectations and enhance its usability (Table 2). Table 2 includes the rubrics of the versions of the tool. The final row which represents the design development and communication skills corresponds the process-oriented version. It can be eliminated for the product-oriented version.

This afore-mentioned summative product-oriented tool is suitable to the evaluation of designs in the competitions, practice and final stages of design in the educational domains. To consider the process-related aspects of design in the educational fields, another version of this model is introduced (Fig. 4). The modified version consists of six blades, and adds a criterion that considers the skills developed during the design development process to the previous five-fold model. These added skills include two indicators, that is, the presentation and the processing skills. The first indicator includes the graphic, verbal, model-making, animation, video production and writing skills needed to communicate ideas. The second is concerned with the progress, processing method, development sequence and teamwork management skills where applicable. The evaluation section of this wheel consists of four concentric hexagons that move inwards from the highest level (A-score) to the lowest (D-score). In the rubrics table, the last row -as illustrated in Table 2- details how to assess this criterion.

5. The tool testing

The third part of this research is experimental, in which the impact of the previous DQE wheel is tested in a design studio. The version that is tested is the product-oriented one because the cases used represent finished building designs that have no information about their processing methods. The following sections detail the experiment ingredients and process. The fourth part analyzes the findings and discusses the results of the model implementation statistically and semantically in the following section.

5.1. The participants

A total of 24 senior-level students participated in the study, among whom 15 students completed the evaluation process and returned the evaluation results ($N = 15$). The students are in their final study year (fifth-year of undergraduate study). They are all female and of Architecture major in Prince Sultan University.

5.2. The procedure

The experimental part of the study consists of two sessions. The first includes an evaluation of a set of design products

Table 2 Supportive design evaluation rubrics.

Criterion	Indicator	1 Inadequate	2 Satisfactory	3 Meets expectations	4 Exceeds expectations
Function	Spatial Layout	Major problems in space design and the spatial refashions	Many spaces lack the right dimensions, proportions and exposure to natural ventilation and views or have weak spatial relationships with other spaces	Minor problems in space design and the spatial relationships	Spaces are well-designed in terms of function, proportion, views, natural light and ventilation.
	Accessibility & Circulation	Major errors in entry points allocation, hierarchy and emphasis, in exit point distribution, in indoor, outdoor, horizontal & vertical circulation elements or in BF solutions.	Frequent errors in entry points selection, hierarchy and emphasis, in exit point distribution, in indoor, outdoor, horizontal & vertical circulation elements or in BF solutions.	Minor errors in entry points allocation, hierarchy and emphasis, in exit point distribution, in indoor, outdoor, horizontal & vertical circulation elements or in BF solutions.	Entry points selection, hierarchy and emphasis are well designed. Exit points are well-located and spaced. All indoor, outdoor, horizontal & vertical circulation and BF elements are well-designed.
Form	Aesthetics	Many appearance &principles of composition problems. Form is not attractive	Minor appearance & compliance with principles of composition problems. Form is repeated in other similar buildings	Pleasant appearance that meets principles of composition with few exceptions	Pleasant appearance that meets principles of composition and provides attractive and unique forms
	Creativity, Novelty & Originality	The design is typical or similar to existing solutions	The design has some aspects of novelty, creativity or originality	The design is mostly creative and original except some aspects	The design is totally creative, novel and original
Context	Site Planning	Major problems in site planning, parking or landscaping schemes	Frequent problems in site planning, parking or landscaping	Minor problems in site planning, parking or landscaping	The building is well-placed on its site, the outdoor, parking & landscape elements are well-designed
	Urban & Environmental Fit	Major problems in the relation between the building & its surroundings and environment	Many minor problems in the relation between the building & its surroundings and environment including the environmental, social & cultural considerations	Minor problems in the relation between the building & its surroundings and environment including the environmental, social & cultural considerations	The relation between the building & its surroundings and environment are well-solved. The environmental, social & cultural considerations are well-incorporated in the design.

(*continued on next page*)

Table 2 (*continued*)

Criterion	Indicator	1 Inadequate	2 Satisfactory	3 Meets expectations	4 Exceeds expectations
Performance	Building systems	Major problems in natural & engineering passive & active building system design & integration	Many minor problems or a major problem in natural & engineering passive & active building system design & integration	Minor problems in natural & engineering passive & active building system design & integration	Natural & engineering passive & active building systems are well designed and integrated
	Human factors & Social Interaction	Major problems in the anthropometrics, ergonomics, human & social factor considerations that may negatively affect the user experience quality	Many minor problems or a major problem in the anthropometrics, ergonomics, human & social factor considerations that may affect the user experience quality	Minor problems in the anthropometrics, ergonomics, human & social factor considerations that may affect the user experience quality	Anthropometrics, ergonomics, human & social factors are well considered and incorporated in the design to improve the end user experience quality
Concept	Identity & Character	Confused character and widely repeated design	Many similar designs exit and for various typologies	The design identity is not directly clear and may not be totally unique	Clear character and unique design
	Semantics (Emotions, symbols, impressions & vibes)	Serious lack of meanings and the design is similar to vacant sculptures	Minor stimulation of positive impressions and connotations	Frequent occurrences of elements that stimulate positive impressions and connotations	Strong stimulation of positive impressions and connotations
Skills	Presentation (Verbal, graphic, animation, model-making & writing). Each in terms of: Correctness, consistency & quality.	Major errors in verbal, graphic or writing presentations or poor presentation of the design.	Frequent minor errors or a major error in verbal, graphic or writing presentations	Minor errors in verbal, graphic or writing presentations, or the media is not effectively presented.	Drawings are correct in terms of scale, graphic representation & consistency between various projections. Verbal, graphic & writing presentations are correct. All media are well-presented.
	Development & Processing	Major errors in design development and processing and poor linkage of the various phases of design processing	Multiple minor errors or a major deficiency in the sequence of design development and processing	Minor errors in in the sequence and linkage of the various phases of design development and processing	Fluent and articulate sequence and linkage of the various phases of design development and processing

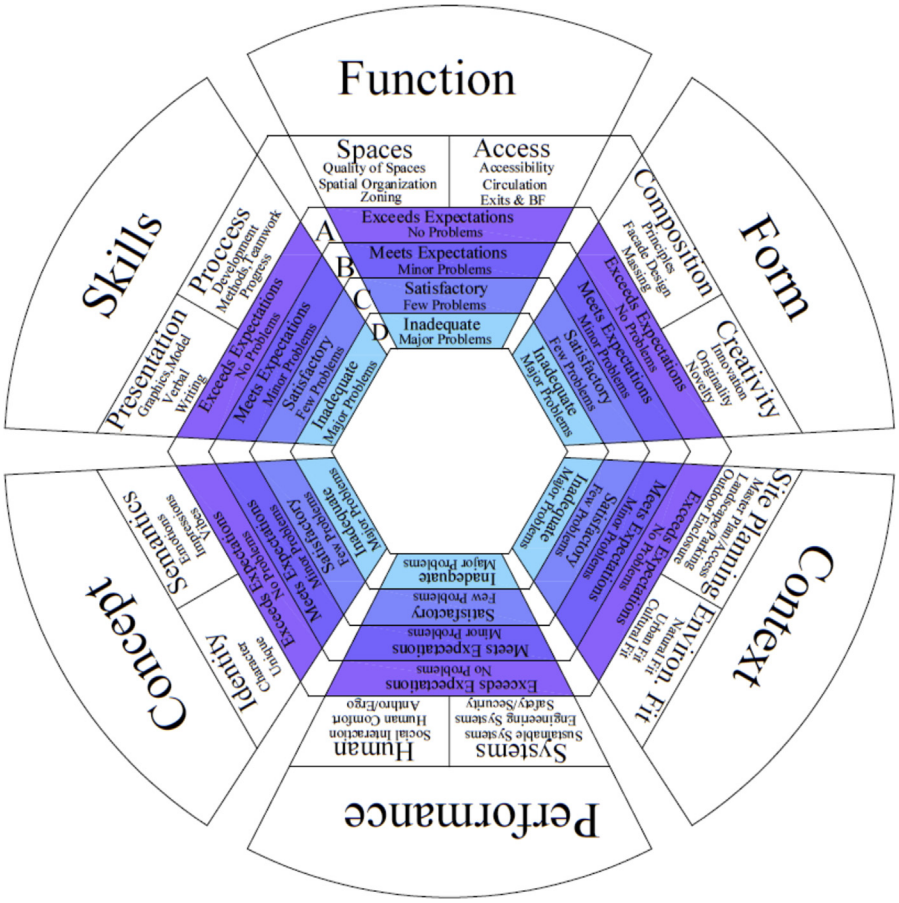


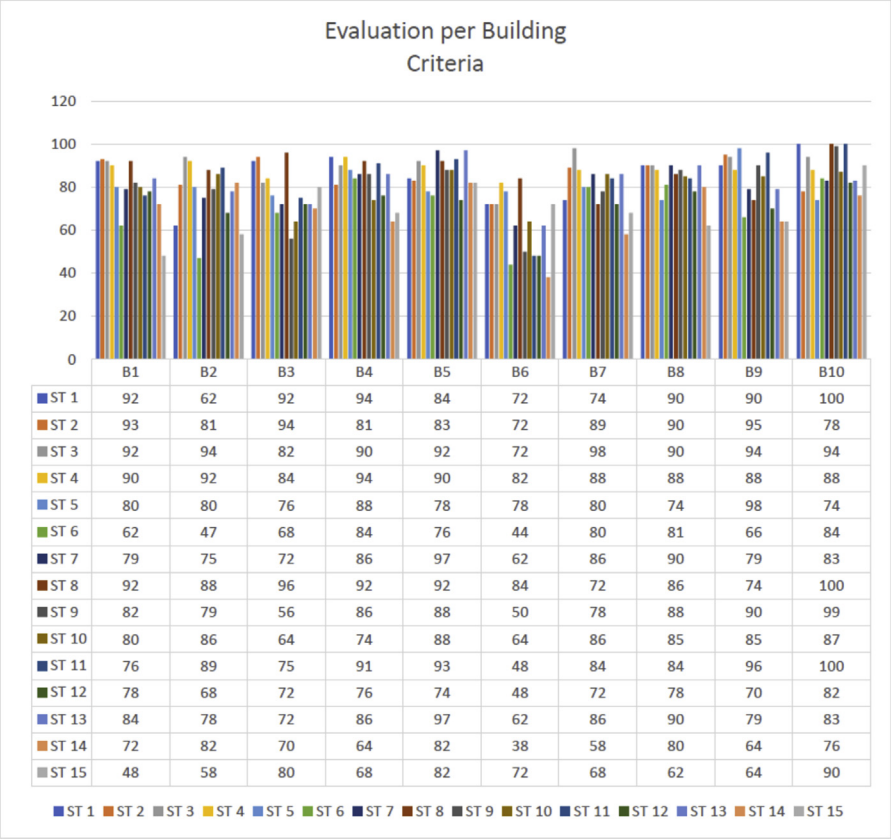
Fig. 4 Deign quality evaluation wheel - Process-oriented version.

Table 3 An example of a criterion-based evaluation table.

Building	Function	Form	Context	Performance	Concept	Total
B1	9	10	10	9	8.5	46.5
B2	9	7	7	8	9.5	40.5
B3	9	10	9.5	9	9.5	47
B4	8	7	9	8	8.5	40.5
B5	7.5	8	8	9	9	41.5
B6	7	7	7	7	8	36
B7	9	8	9	9	9.5	44.5
B8	9	9	9	8	10	45
B9	9	10	10	9.5	9	47.5
B10		10	10	9	10	39

that is conducted based on a table of design criteria only (e.g. Table 3). The second session is conducted in another day and repeats the process of design evaluation of the same design products, but this time it is conducted using the evaluation wheel. Since the design products represent complete precedents, the pentagon version (the product-oriented version) of the DQE wheel is used (Fig. 3). Each session was organized separately. Each participant did the task individually without any information given about the other participants.

In each session, a set of ten building designs was displayed and explained to the participants. These were selected to represent various aspects of design, such as various levels of creativity, harmony with environment, originality, novelty, aesthetics, complexity, and cultural reflectiveness. Each building was represented by drawings and images that express its function (plans and sections), form (exterior three-dimensional shots and elevations) and contexts (site planning and contextual images). For each building, the participants were asked to evaluate its



	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
ST 1	92	62	92	94	84	72	74	90	90	100
ST 2	93	81	94	81	83	72	89	90	95	78
ST 3	92	94	82	90	92	72	98	90	94	94
ST 4	90	92	84	94	90	82	88	88	88	88
ST 5	80	80	76	88	78	78	80	74	98	74
ST 6	62	47	68	84	76	44	80	81	66	84
ST 7	79	75	72	86	97	62	86	90	79	83
ST 8	92	88	96	92	92	84	72	86	74	100
ST 9	82	79	56	86	88	50	78	88	90	99
ST 10	80	86	64	74	88	64	86	85	85	87
ST 11	76	89	75	91	93	48	84	84	96	100
ST 12	78	68	72	76	74	48	72	78	70	82
ST 13	84	78	72	86	97	62	86	90	79	83
ST 14	72	82	70	64	82	38	58	80	64	76
ST 15	48	58	80	68	82	72	68	62	64	90

Fig. 5 Evaluation of the buildings using the criterion-based method.

design quality in terms of its *Functional Solution*, *Form*, *Contextual Fit*, *Performance*, and *Concept*. For each criterion, the participants were required to assign a score out of 10, where 10 indicates a full satisfaction with the criterion fulfillment and 0 indicates an extreme dissatisfaction. The nominal assessment was augmented by a numerical assignment to facilitate the statistical inference of the outcomes. However, no prioritization was added to the criteria to keep the calculations simple. To avoid the mutual influence of each participant on others, the sample members were asked to evaluate the designs individually. After the evaluation sessions, the anonymous responses were collected, translated to computer input and analyzed in terms of their variances, means, modes and deviations.

6. Findings and discussion

The responses that represent the evaluation scores of the various buildings in the sample set consist of two sets, that is, the criterion-based evaluation and the DQE-based evaluation. The main goal of organizing the two sessions is to compare the impacts of using the DQE wheel on the evaluation outcomes with those resulting from using a conventional evaluation method.

The results of using a conventional criterion-based evaluation are illustrated in Fig. 5. Fig. 5 illustrates the total evaluation for each one of the ten buildings according to the given design criteria. The horizontal axis represents the students' scores of the buildings, whereas the vertical axis represents the total score (out of 100) recorded by each student for each building. Hence, the columns illustrate the variety of scores given by all students to each building.

To further represent the scores, a three-dimensional representation of the scores is illustrated in Fig. 6. Each slice represents the scores given by all students to each building.

Similarly, Fig. 7 illustrates the correspondent results as recorded by the students for the same ten buildings using the DQE wheel. The three-dimensional representation of the scores that were recorded using the DQE wheel is illustrated in Fig. 8.

The two sets of evaluations are compared in Table 4. Each row in the table represents a building. Each column illustrates the scores graph for each building using one of the evaluation methods. For each graph in the table, the horizontal X-axis represents the 15 student responses, the vertical Y-axis represents the score given by each student to the building allocated in the table row. As mentioned earlier, each table row addresses a building, and each column addresses an evaluation method.

A statistical summary of the two sets of evaluation results is illustrated in Table 5.

The scores recorded for each design by each participant were analyzed in terms of their statistical means, ranges, nodes and standard deviations. Correlations between the

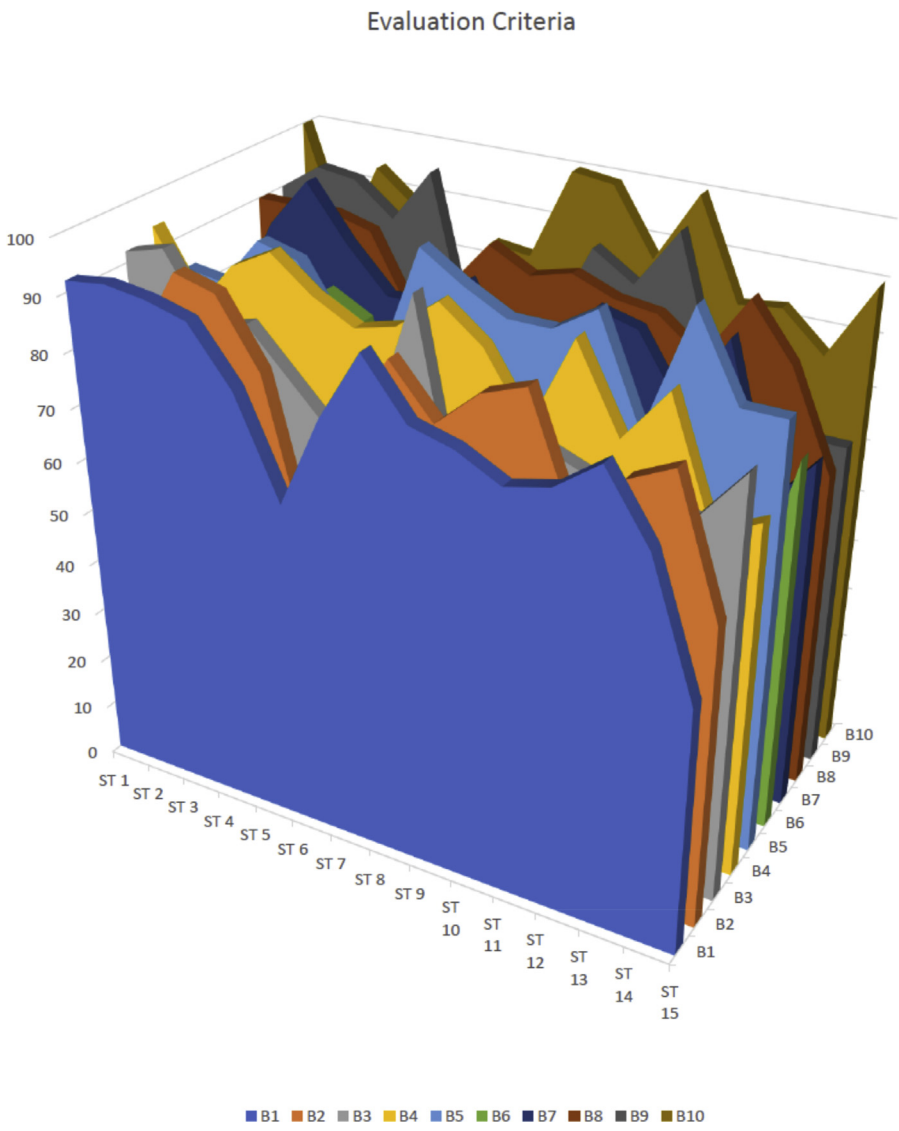


Fig. 6 Three-dimensional representation of the evaluation of the buildings using the criterion-based method.

scores were not studies at this stage. They may be added in a future extension of this research.

As a result of the analysis and comparison of the two evaluation sets of the building designs in the selected sample as conducted by the students using the two methods of evaluation, the following observations can be recorded:

- The responses of the evaluation experiment's participants demonstrate major differences in the evaluation scores of the building designs using the two methods of evaluation. For example, the range of difference in evaluation scores for some buildings reached its extreme of 4.7 steps between one assessor and another on a scale out of 10. This difference was recorded in the evaluation of building B2 using the criterion-based evaluation.
- Among the criteria, the evaluations of concept and form witnessed the highest differences. The least variations were recorded for the function criterion. Hence, the objective judgement of function seems easier than that of form or concept.
- The scores recorded using the DQE wheel witnessed less standard-deviations than those using criteria alone. As such, the wheel seems to contribute to the objective judgement of design products.
- The least standard deviation is recorded in both evaluation methods for building B5 that combines simplicity and creativity.
- The least range - difference between the highest score and the lowest-is recorded for the highly creative buildings (B5 & B8).
- The highest range of scores is recorded for the rigid and sharp buildings (B2 & B6).
- In both evaluation methods, the highest standard deviation is recorded for the building perceived as the least aesthetic (B6).
- The highest average is different in each method, it is assigned for a simple contextual design (B1) in the

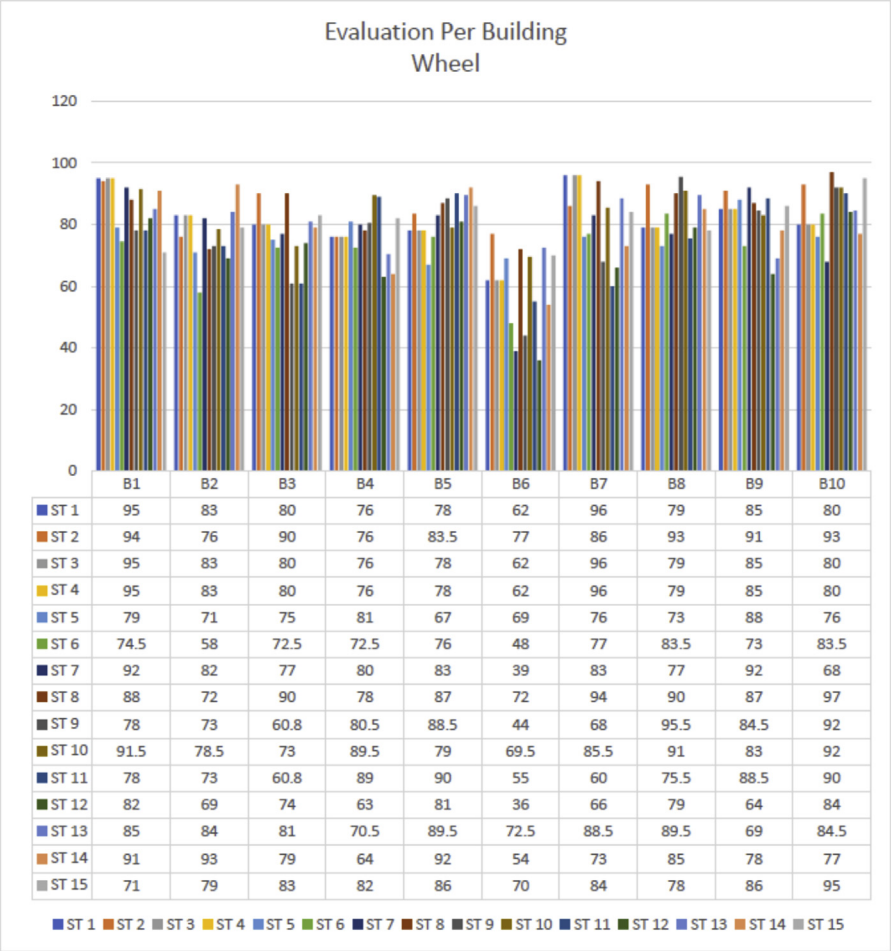


Fig. 7 Evaluation of the buildings using the DQE wheel-based method.

criterion-based evaluation, and for the most imaginative (B10) in the DQE-based evaluation.

- The lowest average is common in both methods. It is assigned for the least aesthetic design (B6).
- Simplicity and minimalism are more valued in the in the DQE-based evaluation method (B1).
- Daring solutions are not enough to guarantee a high score if not combined by aesthetics (B6). In both methods, B6 is assigned the lowest score despite the challenging solution exhibited in its design. However, when challenge is combined with creativity and environmental solution in a building, its design gets high scores in both methods (B5 and B10).
- It is apparent that the more conservative the design is (e.g. B1), the higher average of its scores is using the DQE wheel method. Similarly, the less aesthetically attractive the design is, the lower score it receives (B6).
- The difference in the two sets of evaluations was reduced into up to 2.1 steps in the DQE evaluation. Thus, the objectivity of scores is increased using the wheel tool. This represents a demonstration of achieving the main goal of the wheel and supports it adoption as an additional tool in the design evaluation field.
- The emotional impact that is related to the cultural ingredient in design is more valued in the criterion-based evaluation (B4) than in the DQE-based evaluation. However, when the cultural heritage-related ingredient is combined with additional aesthetic values, it is valued equally in both methods (B9). It was noticed that the buildings that represented a style related to the cultural background of the participants were scored highly although they were not normally considered creative or challenging. It seems that buildings that evoke an emotion such as the nostalgia with associations to the past of the participant score higher than what their mere designs deserve. This interesting observation supports the relatively new orientation of "affective engineering" which emphasizes the intangible elements of emotions and attractions as major factors in design quality perception and consequently in marketing potentials (e.g. Hanamachi, 2010; Schütte et al., 2008; Singh, 2013; Watada et al., 2014). It also supports the cultural tendencies impact on design processing as suggested by Van Dooren et al. (2014). This interesting finding can be further addressed in a future extension of this research.
- The study has two limitations. The first is related to the small size of its sample. The second is that it did not consider the element of repetition and its influence on the final results as pointed out by Coughlan and

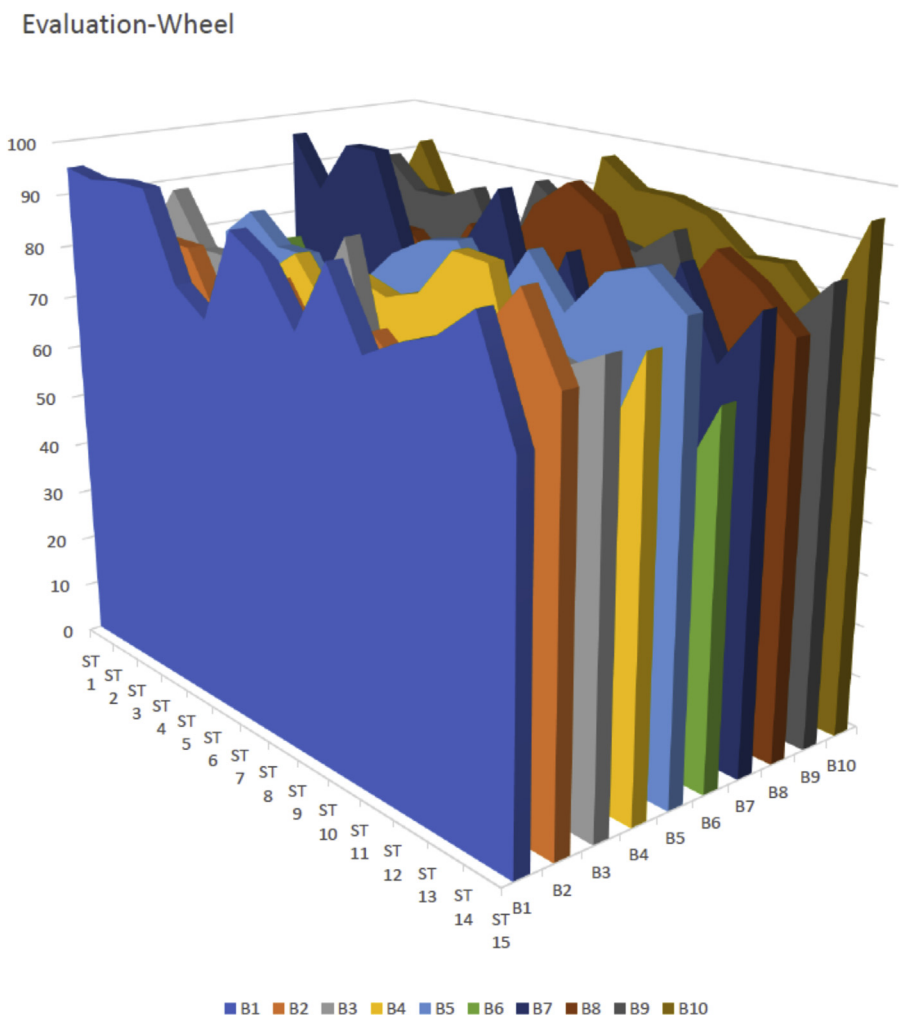


Fig. 8 Three-dimensional representation of the evaluation of the buildings using the DQE wheel-based method.

Mashman (1999) earlier in the literature review of this study.

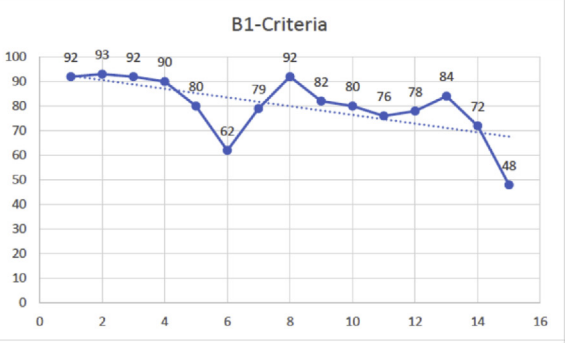
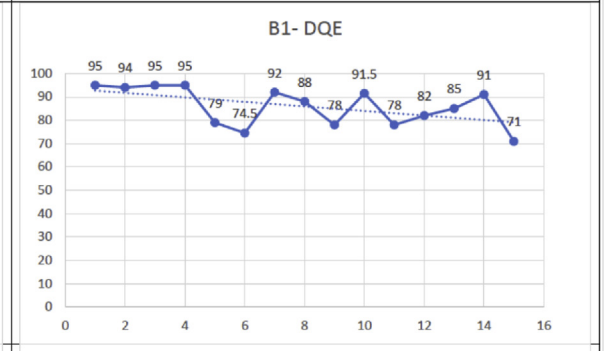
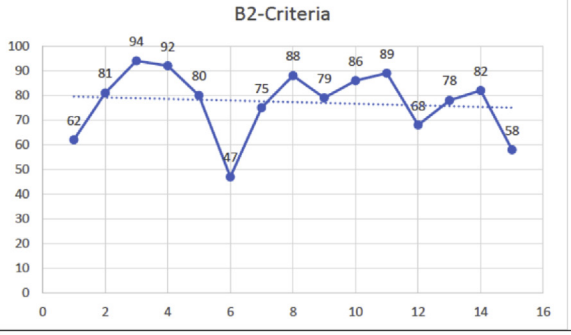
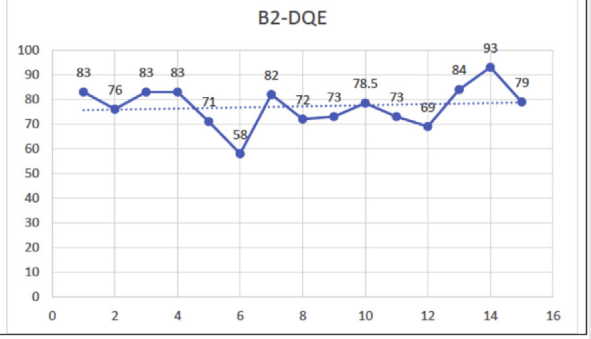
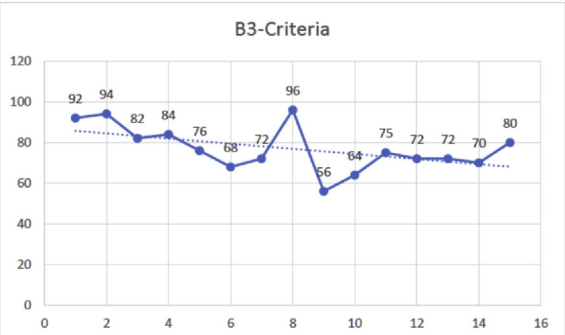
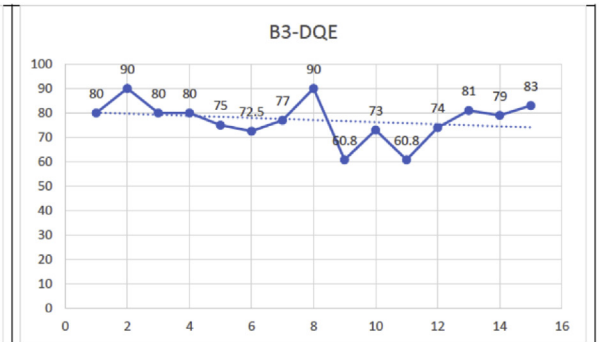
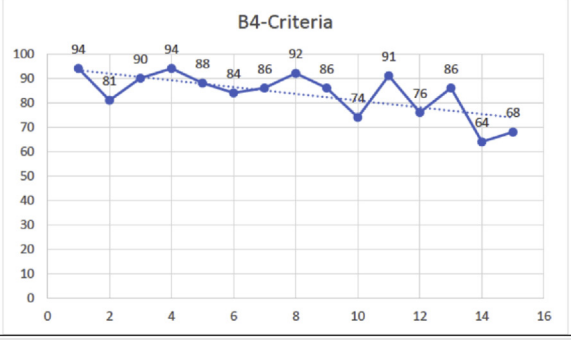
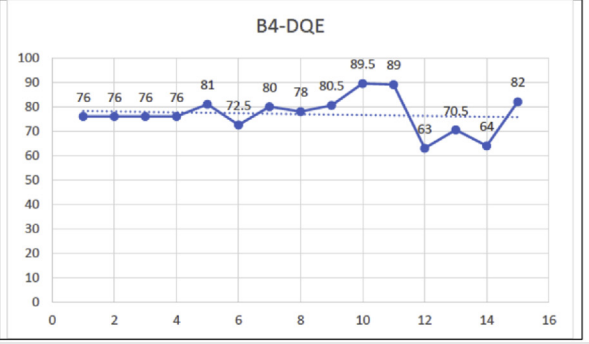
As a result of the previous observations, the role of the tool in the reduction of discrepancies of the scores is clear. Notably, despite the small sample size of this experiment, its outcomes are still enough for supporting the validity of its presented findings at this stage. Typically, the big difference in the scores of various evaluators of design products seems to express a lack of objectivity in the evaluators' side and confusion and disappointment in the designers' side. Even a slight reduction of such a difference is a significant achievement for the presented tool. The succinct graphic representation of the tool and its scales seem to help the assessors evaluate the products and guide them throughout the evaluation process. Future extensions of this research may address larger sample sizes and repeat similar experiments to increase the validity and reliability of the findings.

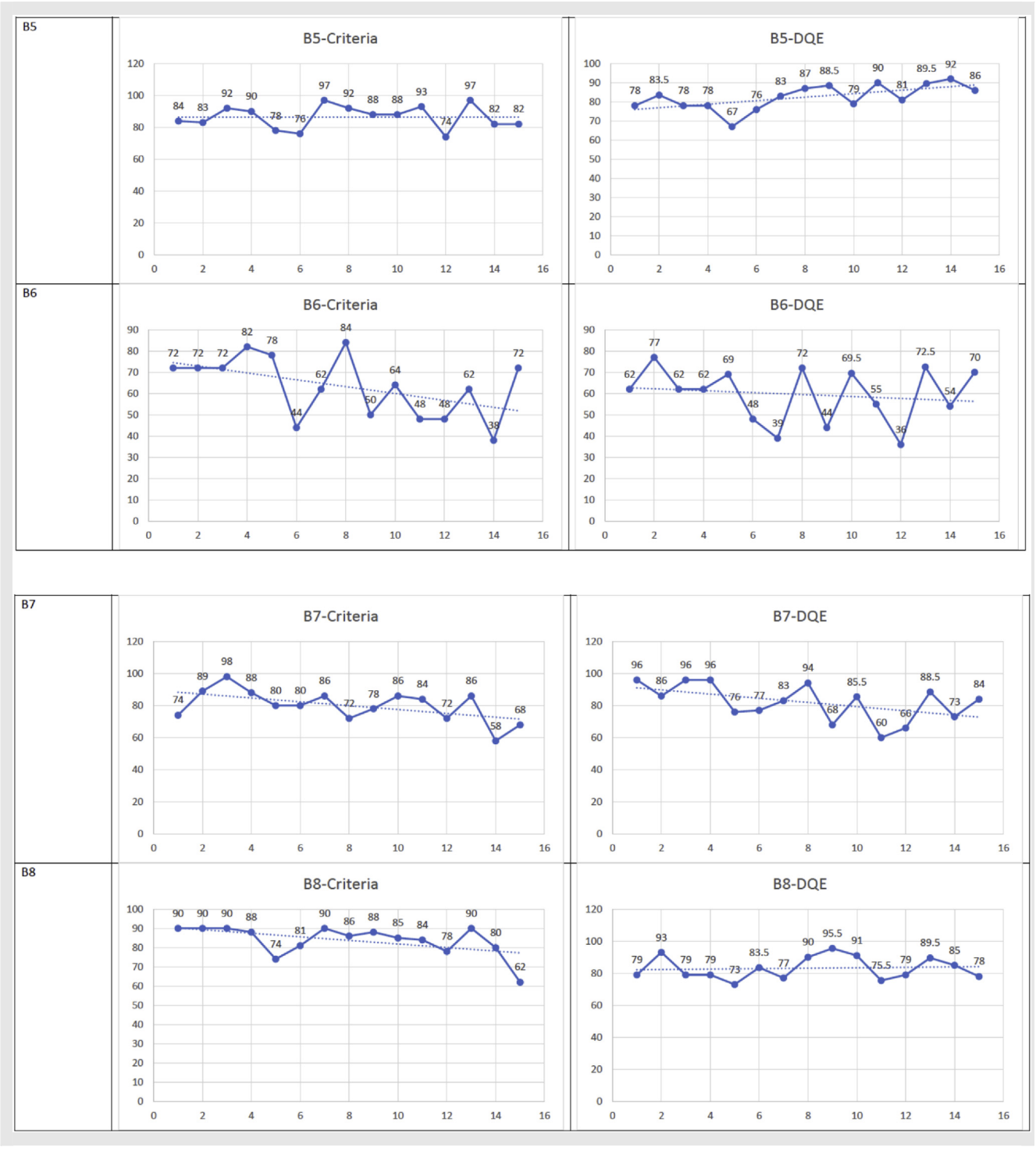
7. Conclusion

Representing a hybrid combination of rational and normative ingredients, architectural design is one of the most difficult subjects to evaluate. In this study, which attempts to contribute to knowledge in this area, a novel design quality evaluation tool is developed and presented. The tool is called the design quality evaluation (DQE) wheel. In addition to the derivative part in which the tool is developed, an explorative part of the study commences this research to establish its knowledgebase. In addition, an experimental and analytical parts are used to test the tool and analyze its implementation implications. Two methods of evaluation are conducted to evaluate a set of building designs. These include the Criterion-Based Evaluation and the DQE Wheel-Based Evaluation methods.

The design quality evaluation tool is introduced as a graphic user-friendly wheel to facilitate evaluation and make results more consistent. The tool combines a set of design evaluation criteria, quality indicators and an incremental scale to incorporate detailed assessment rubrics. To test its applicability, ten building designs were displayed to a group of senior-level students of Architecture major in a higher education studio setting. The students were asked to evaluate the ten buildings using two methods of evaluation. The first is a conventional table-formed criterion-based evaluation, and the second is using the DQE wheel. Their scoring records are analyzed and

Table 4 Comparison of the two evaluation scores.

Building	Criterion-Based Evaluation	DQE Wheel-Based Evaluation
B1		
B2		
B3		
B4		


B5
B5-Criteria
84 83 92 90 78 76 97 92 88 88 93 74 97 82 82
B5-DQE
78 83.5 78 78 67 76 83 87 88.5 79 90 81 89.5 92 86
B6
B6-Criteria
72 72 72 82 78 44 62 84 50 64 48 48 62 38 72
B6-DQE
62 77 62 62 69 48 39 72 44 69.5 55 36 72.5 54 70
B7
B7-Criteria
74 89 98 88 80 80 86 72 78 86 84 72 86 58 68
B7-DQE
96 86 96 96 76 77 83 94 68 85.5 60 66 88.5 73 84
B8
B8-Criteria
90 90 90 88 74 81 90 86 88 85 84 78 90 80 62
B8-DQE
79 93 79 79 73 83.5 77 90 95.5 91 75.5 79 89.5 85 78

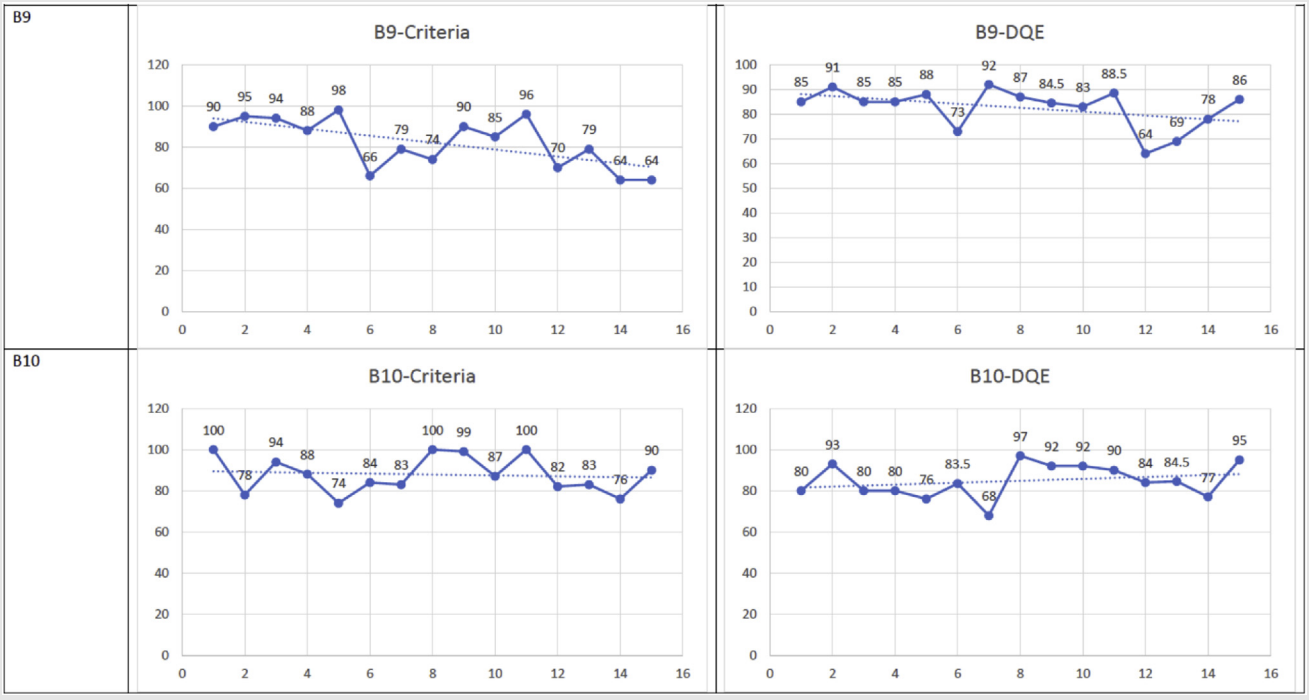


Table 5 A statistical summary of the two sets of evaluation results.

Building ↓	Criterion-Based Evaluation				DQE Wheel-Based Evaluation				Results	
	Average	Mode	Range	Standard Deviation	Average	Mode	Range	Standard Deviation	Steps	Std Dev Diff
B1	80.0	92	45	12.40	85.9	95	24	8.25	2.1	4.15
B2	77.3	N/A	47	13.32	77.2	83	35	8.29	1.2	5.03
B3	76.9	72	40	11.27	77.1	80	29.2	8.43	1.08	2.84
B4	83.6	86	30	9.27	76.9	76	26.5	7.54	0.35	1.73
B5	86.4	92	23	7.26	82.4	78	25	6.67	−0.2	0.59
B6	63.2	72	46	14.54	59.5	62	41	12.95	0.5	1.59
B7	79.9	86	40	9.96	81.9	96	36	11.60	0.4	−1.64
B8	83.7	90	28	7.84	83.1	79	22.5	7.05	0.55	0.79
B9	82.1	90	34	12.15	82.6	85	28	8.10	0.6	4.05
B10	87.9	100	26	9.01	84.8	80	29	8.21	−0.3	0.8

compared to find out their differences, commonalities, variations, associations, and implications. Some interesting findings resulted from the analysis of this experiment. These include that the less aesthetic a design is, the lower its evaluation is in both types of evaluation. Another unexpected finding was the influence of the affective dimensions on evaluation. An example of these is the impact of the cultural attachment and the traditional connotations in design on its evaluation. These affective dimensions seem to add value and increase the evaluators' satisfaction with the product design. As a result of the tool application, it seems that using the DQE wheel contributes to making the differences in scores between the different assessors less, and consequently the DQE wheel seems to improve the objectivity of evaluation. This result supports Lindström (2006) proposal that detailed rubrics can reduce the differences in evaluation scores. In this study, these differences were reduced to 2.4 steps (as opposed to 4.5 steps using the criterion-based evaluation). Thus, the hypothesis that the method of the evaluation influences the final judgement of design products and their outcomes is validated through the score analyses and comparisons in this research. However, the study has two limitations. The first is related to the small size of its sample. The second is that it did not consider the element of repetition and its influence on the final results.

Multiple extensions may continue and build up on this research. One of these will be its duplication in various settings to validate its results. Furthermore, it may be duplicated with the same participants and same evaluation method to check Coughlan and Mashman (1999) hypothesis that repetition of evaluation may change its results. Moreover, since the participants in this study were all females, another extension will be to repeat the experiment

for all male evaluators to study the impact of gender on the evaluation result. Another extension of this study is to test the DQE tool in design studios with larger sample sizes to validate its applicability and overcome its current limitations. In addition, the wheel's structure may be modified to fit other engineering design fields. It may also be developed to a manual tool where operative rotational dials may be added in layered sheets to enable evaluators to get the final scores easily. Alternatively, the wheel may be developed to a touchable computerized version where the total of the indicators' scores can be calculated automatically.

References

Allen, M.J., 2004. Assessing Academic Programs in Higher Education. Jossey—Bass, San Francisco.

Armstrong, C.L., 1994. Designing Assessment in Art. National Art Education Association, Reston, VA.

Ball, L.J., Ormerod, T.C., 1995. Structured and opportunistic processing in design: a critical discussion. Int. J. Hum. Comput. Stud. 43, 131—151.

Ball, L.J., Ormerod, T.C., Morley, N.J., 2004. Spontaneous analogising in engineering design: a comparative analysis of experts and novices. Des. Stud. 25 (5), 495—508.

Bramley, P., Newby, A.C., 1984. The evaluation of training Part I: clarifying the concept. J. Eur. Ind. Train. 8 (6), 10—16.

Buchanan, R., 1992. Wicked problems in design thinking. Des. Issues 3 (2), 5—21.

Choy, R., Burke, N., 2006. Quality specifications for clients. In: Clients Driving Innovation: Moving Ideas into Practice. Cooperative Research Centre (CRC) for Construction Innovation.

Chyung, S.Y., 2008. Foundations of Instructional Performance Technology. HRD Press Inc, Amherst, MA.

Clancey, W.J., 1985. Heuristic classification. Artif. Intell. 27 (3), 289—350.

Coughlan, P., Mashman, R., 1999. Once is not enough: repeated exposure to and aesthetic evaluation of an automobile design prototype. Des. Stud. 20 (6), 553—563.

Cowan, J., 2000. Evaluation and feedback in architectural education. In: Nicol, D., Pilling, S. (Eds.), Changing Architectural Education. Taylor and Francis Publications, Amherst, MA.

Craig, R.L., 1996. The ASTD Training: Development Handbook. McGraw—Hill, New York.

Cross, N., 2006. Designerly Ways of Knowing. Birkhauser, Basel, Switzerland.

Demirkan, H., Afacan, Y., 2012. Assessing creativity in design education: analysis of the creativity factors in the first—year design studio. Des. Stud. 33 (3), 262—278.

Dijks, Monique A., Brummer, Leonie, Kostons, Danny, 2018. The anonymous reviewer: the relationship between perceived expertise and the perceptions of peer feedback in higher education. Assess Eval. High Educ. 43 (8), 1258—1271.

Dorst, C.H., Cross, N.G., 2001. Creativity in the design process: Co—evolution of problem—solution. Des. Stud. 22 (5), 425—437.

Eilouti, B., 2007. A problem-based learning project for computer-supported architectural design pedagogy. Art Des. Commun. High Educ. 5 (3), 197—212.

Eilouti, Buthayna, 2009. Design knowledge recycling using precedent—based analysis and synthesis models. Des. Stud. 30 (4), 340—368.

Eilouti, Buthayna, 2018a. Concept as the DNA for morphogenesis: a case study of contemporary architecture. In: D'Uva, Domenico (Ed.), Handbook of Research on Form and Morphogenesis in Modern Architectural Contexts. 283—309. IGI Global, Hershey, PA.

Eilouti, Buthayna, 2018b. Concept evolution in architectural design: an octonary framework. Front. Arch. Res. 7 (2), 180—196.

Eilouti, Buthayna, 2018c. Systematic concept derivation and translation in engineering design. In: Proceedings of the 22nd World Multi—Conference on Systemics, Cybernetics and Informatics, vol. III. WMSCI 2018, Orlando, Florida, pp. 188—193.

Eilouti, Buthayna, 2018d. Concept as a Bridge between Abstraction and Concretization in Design Knowledge Visualization. In: IV2018 — 22nd International Conference Information Visualisation. 10 — 13 July 2018. University of Salerno, Salerno, Italy.

Eilouti, Buthayna, 2019. Models for concept derivation and materialization in design management. Int. J. Art Cult. Des. Technol. 8 (1), 51—67.

Farrell, R., Hooker, C., 2013. Design, science and wicked problems. Des. Stud. 34 (6), 681—705.

Fisher, R., Cavanagh, J., Bowles, A., 2011. Assisting transition to university: using assessment as a formative learning tool. Assess Eval. High Educ. 36 (2), 225—237.

Fu, K.K., Yang, M.C., Wood, K.L., 2015. Design principles: the Foundation of design. In: The ASME International Design Engineering Technical Conferences, Boston, MA.

Gann, D.M., Salter, A.J., Whyte, J.K., 2003. Design quality indicator as a tool for thinking. Build. Res. Inf. 31 (5), 318—333.

Georgiev, G.V., Nagai, Y., Taura, T., 2010. A method for the evaluation of meaning structures and its application in conceptual design. J. Des. Res. 8 (3), 214—234.

Gielen, M., De Wever, B., 2015. Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. J. Comput. Assist. Learn. 31 (5), 435—449.

Gilbert, T., 1998. A leisurely look at worthy performance. In: Woods, Gortada (Ed.), The 1998 ASTD Training and Performance Yearbook. McGraw-Hill, New York.

Girod, M., Elliott, A.C., Wright, I.C., 2000. Decision—making and design concept selection. In: Proceedings of the Engineering Design Conference. Brunel University, Middlesex, pp. 659—666, 2000.

Glaser, R., 1963. Instructional technology and the measurement of learning outcomes: some questions. Am. Psychol. 18 (8), 519—521.

Goldstein, D.G., Gigerenzer, G., Hogarth, R.M., Kacelnik, A., Kareev, Y., Klein, G., et al., 2001. Why and how do simple heuristics work? In: Gigerenzer, G., Selten, R. (Eds.), Bounded Rationality: the Adaptive Toolbox. 173—190. MIT Press, Cambridge.

Guindon, R., 1990. Designing the design process: exploiting opportunistic thoughts. Hum. Comput. Interact. 5, 305—344.

Hanamachi, Mitsuo (Ed.), 2010. Kansei/Affective Engineering. CRC Press, New York.

Hickman, R., 2007. Wipped—fancying and other vices: Re—evaluating assessment in art and design. In: Raymend, T. (Ed.), The Problem of Assessment in Art and Design. Intellect Books, Bristol.

Holmes, N., 2015. Student perceptions of their learning and engagement in response to the use of a continuous E—assessment in an undergraduate module. Assess Eval. High Educ. 40 (1), 1—14.

Hubka, V., 1992. Design for quality and design methodology. J. Eng. Des. 3 (1), 5—15.

Hurteau, M., Houle, S., Mongiat, S., 2009. How legitimate and justified are judgments in program evaluation? Evaluation 15 (3), 307—319.

Huxham, Mark, Scoles, Jenny, Green, Ursula, Purves, Samantha, Welsh, Zoe, Gray, Andrew, 2017. Observation has set in: comparing students and peers as reviewers of teaching. Assess Eval. High Educ. 42 (6), 887—899.

Jansson, D.G., Smith, S.M., 1991. Design fixation. Des. Stud. 12 (1), 3—11.
Kirkpatrick, D.L., 1975. In: Kirkpatrick (Ed.), Techniques for Evaluating Training Programs. Evaluating Training Programs. ASTD, Alexandria, VA.
Kite, James, Phongsavan, Philayrath, 2017. Evaluating standards—based assessment rubrics in a postgraduate public health subject. Assess Eval. High Educ. 42 (6), 837—849.
Lera, S.G., 1981. Empirical and theoretical studies of design judgement: a review. Des. Stud. 2 (1), 19—26.
Lindström, L., 2006. Creativity: what is it? can you assess it? can it be taught? J. Art Des. Educ. 25 (1), 53—66.
López-Pastor, Victor, Sicilia-Camacho, Alvaro, 2017. Formative and shared assessment in higher education. Lessons learned and challenges for the future. Assess Eval. High Educ. 42 (1), 77—97.
Lowe, J.B., 1972. The assessment of students' architectural design drawings. Arch. Res. Teach. 2 (2), 96—104.
Marden, N.Y., Ulman, L.G., Wilson, F.S., Velan, G.M., 2013. Online feedback assessments in physiology: effects on students' learning experiences and outcomes. Adv. Physiol. Educ. 37 (2), 192—200.
Markus, H., Ruvolo, A., 1990. Possible selves: personalized representations of goals. In: Pervin, L.A. (Ed.), Goal Concepts in Psychology. Lawrence Erlbaum, Hillsdale, NJ, pp. 211—241.
Murphy, Keith, Ivarsson, Jonas, Lymer, Gustav, 2012. Embodied reasoning in architectural critique. Des. Stud. 33 (6), 530—556.
Nicol, D., Thomson, A., Breslin, C., 2014. Rethinking feedback practices in higher education: a peer review perspective. Assess Eval. High Educ. 39 (1), 102—122.
Nikander, Jan B., Liikkanen, Lassi A., 2014. The preference effect in design concept evaluation. Des. Stud. 35 (5), 473—499.
Palmer, E., Devitt, P., 2014. The assessment of a structured online formative assessment program: a randomised controlled trial. BMC Med. Educ. 14 (1), 8.
Rayment, T. (Ed.), 2007. The Problem of Assessment in Art and Design. Intellect Books, Bristol.
Reeve, J., Peerbhoy, D., 2007. Evaluating the evaluation: understanding the utility and limitations of evaluation as a tool for organizational learning. Health Educ. J. 66 (2), 120—131.
Ross, P.H., Ellipse, M.W., Freeman, H.E., 2004. Evaluation: A Systematic Approach, seventh ed. Sage, Thousand Oaks.
Saaty, T.L., 1977. A scaling method for priorities in hierarchical structures. J. Math. Psychol. 15 (3), 234—281.
Saaty, T.L., Erdener, E., 1979. A new approach to performance measurement: the analytic hierarchy process. Des. Meth. Theor. 113 (2), 64—72.
Saettler, P., 1990. The Evolution of American Educational Technology. Libraries Unlimited Inc, Englewood, Colorado.
Schütte, Simon, Krus, Petter, Eklund, Jörgen, 2008. Integration of affective engineering in product development processes. In: 11th QMOD Conference: Quality Management and Organizational Development Attaining Sustainability from Organizational Excellence to Sustainable Excellence; in Helsingborg—Sweden, pp. 651—660.
Scriven, M., 1967. The methodology of evaluation. In: Tyler, R.W., Gagne, R.M., Scriven, M. (Eds.), Perspectives of Curriculum Evaluation. Rand McNally, Chicago, IL, pp. 39—83.
Shute, V.J., 2008. Focus on formative feedback. Rev. Educ. Res. 78 (1), 153—189.
Singh, Amitoj, 2013. Managing Emotion in Design Innovation. CRC Press, New York.
Snowball, J.D., Mostert, M., 2013. Dancing with the devil: formative peer assessment and academic performance. High. Educ. Res. Dev. 32 (4), 646—659.
Suskie, Linda, 2004. Assessing Student Learning. Anker, Bolton, MA.
Thompson, James, Houston, Don, Dansie, Kathryn, Rayner, Timothy, Pointon, Timothy, Pope, Simon, Cayetano, Anthea, Mitchell, Brad, Grantham, Hugh, 2017. Student and tutor consensus: a partnership in assessment for learning. Assess Eval. High Educ. 42 (6), 942—952.
Thomson, S., 2007. Sharing understanding of assessment criteria in design project tutorials: some observations of, and implications for, practice. IDEA Inter. Des. Inter. Arch. Educ. Assoc. J. 38 (5), 38—50.
Tromp, Nynke, Hekkert, Paul, 2016. Assessing methods for effect—driven design: evaluation of a social design method. Des. Stud. 43 (2), 24—47.
Ullman, D.G., Herling, D., Sinton, A., 1996. Analysis of protocol data to identify product information evolution and decision-making process. In: Cross, N., Christiaans, H., Doorst, K. (Eds.), Analysing Design Activity. Wiley, Chichester, pp. 169—185.
Van der Voordt, D.J.M., Van Wegen, H.B.R., 2005. Architecture in Use: an Introduction to the Programming. Design and Evaluation of Buildings. Architectural Press, Oxford.
Van Dooren, E., Asselbergs, T., Boshuizen, E., Van Merrienboer, J., Van Dorst, M., 2014. Making explicit in design education: generic elements in the design process. Int. J. Technol. Des. Educ. 24 (1), 53—71.
Volker, Leentje, Lauche, Kristina, Heintz, John L., de Jonge, Hans, 2008. Deciding about design quality: design perception during a European tendering procedure. Des. Stud. 29 (4), 387—409.
Watada, J., Shiizuka, H., Lee, K.—P., Otani, T., Lim, C.—P. (Eds.), 2014. Industrial Applications of Affective Engineering. Springer International Publishing, Switzerland.
Webster, H., 2006. Power, freedom and resistance: excavating the design jury. J. Art Des. Educ. 25 (3), 286—296.