The Fourth International Conference on Design Creativity (4th ICDC)
Atlanta, GA, November 2nd-4th, 2016

DEVELOPING METRICS TO ASSESS TECHNOLOGY-ENABLED CREATIVE CO-DESIGN SESSIONS

J. A. O'Hare[1], E. A. Dekoninck[1] , M. Mombeshora[1], and G. Cascini[2]

[1]Dept. Mechanical Engineering, University of Bath, U.K.
[2] Dipartimento di Meccanica, Politecnico di Milano, Italy.

Abstract: Users, customers and clients are playing an increasingly important role in the design process as part of recent trends towards 'co-creation' and 'co-design'. Co-design sessions seem to be an important part of the co-design process but can be challenging for designers, for whom collaborating with people not trained in design is a new activity. There are a number of tools and approaches that might help overcome such challenges but to evaluate these tools we need some performance metrics for co-design sessions. This paper reports on the development of a novel set of quantitative metrics that can be used to evaluate the performance and effectiveness of co-design sessions. Building on existing metrics found in the academic literature, the new metrics cover activities such as idea generation, idea filtering, and progress on project tasks. Future work will see these metrics applied to real-life case studies as part of the SPARK project (www.spark-project.net).

Keywords: *metrics, augmented reality, co-design, co-creation*

1. Introduction

It has long been understood that successful product development requires a clear understanding of user needs and requirements. What has changed in recent decades is the role of the user within the design process. According to Sanders & Stappers (2008) the role of the user has changed from being the 'subject' of study (through approaches such as usability testing) to being a 'partner' in the process. This transition has been driven by the belief that if designers can engage with users in a more profound, expansive and cooperative way during the design process then they will increase their chances of delivering a final product that effectively fulfils users requirements and is commercially successful. Recently the European Commission (EU, 2014) recognized that customers are having a more and more active role in innovation, showing that the standard, and now obsolete, seller/buyer relationship has undergone a transition to a business-customer relationship where the customers are not just goods consumers, but co-creators and co-designers.

According to Sanders & Stappers (2008), the term 'co-creation' can be used to describe "…any act of collective creativity, i.e. creativity that is shared by two or more people". From this broad definition a number of different types of co-creation can be identified, such as ‘collaboration’, ‘tinkering’, ‘co-design’ and ‘submitting’ (O’Hern & Rindfleisch, 2008). Within this paper, the focus is on ‘co-design’, which is defined as “the creativity of designers and people not trained in design working together in the design development process.” (Sanders and Stappers, 2008). We use the term ‘non-designer’ in the rest of the paper to refer to participants of a co-design session that are not trained in design, for example users, customers or clients.

Responding to this trend, the design research community has begun to propose tools and approaches that are aimed at improving the effectiveness and efficiency of co-design activities (Sanders and Stappers, 2014). The design of these tools is important as they can determine the chance of success in the design activitiy and the users’ willingness to participate in future co-design activities (Füller et al., 2009).

Within the topic of tools to support co-creation, augmented reality and virtual reality have been noted as technologies that seem very relevant. This potential serves as the starting point for the the recently launched SPARK project (www.spark-project.net), the aim of which is to understand how Spatial Augmented Reality (SAR) technology can be used to support the co-design process. A particular focus of the SPARK project will be on the ‘co-design session’, which we define as ‘a pre-arranged session that involves designers and people not trained in design working together in the design development process’. This definition allows for a broad range of activities such as idea generation, reviewing and selecting concepts, creating plans for testing, etc. This is reflected in Figure 1, below, which shows a (non-exhaustive) spectrum of types of co-design sessions, ranging from ‘100% Creative’ to ‘100% Selection’. In the middle we define ‘Creative review’ as a type of co-design session in which pre-existing ideas are reviewed, but which also involves some new idea generation.

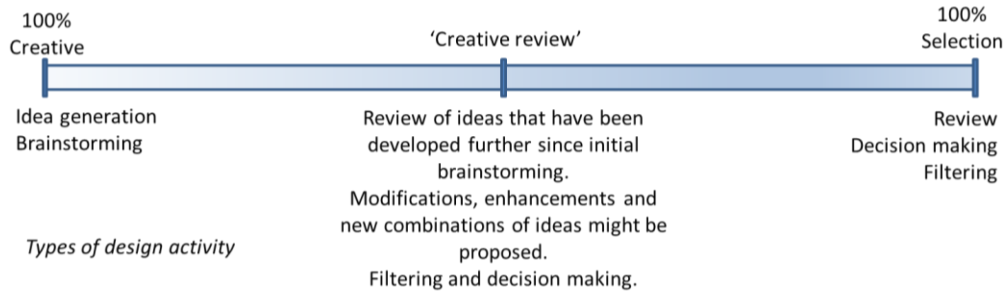

Figure 1. The spectrum of types of co-design session.

Some of the features of co-design sessions present new challenges for practitioners trying to facilitate this type of session, Taffe’s (2015) examples include: lack of shared design vocabulary; and designers behaving as the sole creative experts. Lee (2008) reflected on co-designing and indentified several challenges for professional practice including: some design skills cannot be transferred to non-designers such as aesthetics; and the mixed terminology used in the field.

Co-design sessions therefore seem to be important and potentially challenging activities within the design process. The SPARK project aims to address some of the challenges of co-design sessions and therefore a key starting point for this research is understanding how the performance and effectiveness of co-design sessions can be measured. Without some measures of performance we cannot be sure whether any proposed Spatial Augmented Reality (SAR) technology enhances the performance of co-design sessions or not. The main aim of this paper is therefore to propose a number of performance metrics for co-design sessions.

2. Methodology

The metrics were developed through a four-step process. These steps were as follows:

Step 1 – Identification of the main activities that take place during a co-design session

An initial list of the typical activities that take place within a co-design session was generated. The main activities identified were: ‘idea generation’, ‘enhancement or modification of existing ideas’, ‘evaluation and filtering of ideas’ and ‘identifying and completing other types of design task’. This list of activities was used to guide the literature review in step 2.

Step 2 - Literature review on metrics used for co-design sessions

The initial literature review searched within the product design and development literature for suitable metrics for the co-design activities from Step 1. The literature search was extended beyond the field of product design and development, into more general management domains, such as organisational management.

Step 3 – Reflection on the suitability of the metrics identified from the literature

Here we considered whether the metrics would be practical to apply within the context of the SPARK project (summarised in section 3).

Step 4 – Selection and development of metrics

Where possible, an existing metric was selected for use within the SPARK project. In the cases where no suitable metric could be identified, new metrics were proposed by the SPARK consortium.

3. Review of literature on metrics for co-creation sessions

This section presents the literature review, dealing first with the metrics of effectiveness of idea generation and then with the metrics for idea evaluation, filtering and progress on tasks.

3.1. Literature on the effectiveness of idea generation

The effectiveness of idea generation activities can be evaluated by methods that can be broadly categorised as being either process- or outcome-based (Shah et al, 2000). In essence, process-based approaches observe the creative cognitive processes of creative thinking; where creative thinking can be defined as ‘… the forming of associative elements into new combinations which either meet specified requirements or are in some way useful’ (Mednick, 1962). Process-based evaluation approaches tend to be subjective - as no commonly agreed upon techniques to conduct them exists - and time-consuming (Verhaegen et al, 2011). Also, it is difficult to establish firm causal links between the occurrence of a cognitive process and the effectiveness of an idea generation tool/technique (Shah et al, 2003).

Outcome-based approaches have gained in popularity in recent years as they address some of these concerns by focusing on the (external, observable) ideas generated rather than the (internal, unobservable) process by which they are generated. With this type of approach, an idea generation tool/technique can then be considered to be effective if its use leads to “good” ideas. The metrics developed for outcome-based approaches therefore attempt to relate aspects of the generated design ideas to the effectiveness of the applied idea generation method (Verhaegen et al, 2011).

The seminal work by Shah et al (2000, 2003) on outcome-based measures defined four metrics that are intended to assess the effectiveness of design space expansion and exploration. The metrics relate to the quantity, quality, variety and novelty of ideas generated and have been used in numerous studies in the field, with several authors proposing refinements to the original metrics (Nelson et al, 2009; Verhaegen et al, 2011; Gosnell and Miller, 2015).

3.2. Literature on the effectiveness of idea evaluation, filtering and progress on tasks

As well as idea generation, co-design sessions can be used for other tasks such as idea evaluation, concept filtering and making important decisions about the design of the product. These types of activity are often part of what is known as a ‘design review’ session. While most textbooks on product development do mention the topic of design reviews as a vital activity within the design process, literature on design reviews is somewhat limited. Wetmore at al. (2010) however, have looked at the effectiveness of design reviews. Their work investigates the effects of the information shared in advance of the session. They measured review effectiveness by looking at the number, and nature of, the design problems that the teams managed to identify. Ostergaard et al.(2003) studied design reviews with a focus on risk: their metric for design review effectiveness is therefore also a sum of the problems identified by the team involved. Hannah et al. (2012) showed that, in design reviews, confidence in judgements is affected by the nature of the design representations used - from sketches to high -fidelity prototypes. Boujut & Blanco (2003) present a detailed anylysis of a typical design review in industry and discuss the role that design representations, and the objects created, play. Within this literature, the various authors emphasise different purposes for design reviews, and therefore measures of effectivity are somewhat disparate. Because our aim was to be able to measure the effectiveness across the whole spectrum of co-design sessions (shown in figure 1), we decided in stead to look beyond product development to management literature for more generic insights on meeting effectiveness.

Comments on meeting effectiveness typically reflect four main themes: people, meeting organisation, meeting activities and meeting outcomes (Geimer et al, 2012). The effectiveness of meetings is often evaluated by analysing the meeting activities as well as factors such as: goal attainment, decision satisfaction, task performance, relationship quality etc. (e.g. Leach et al, 2009). These assessments of meeting effectiveness are based on participant perceptions, and are subjective in nature. In their white paper, Lucid Meetings (2015) offer an alternative approach to evaluating meeting effectiveness that emphasises collecting quantitative data about meetings in order to provide objective data points that can be correlated with meeting quality. They suggest that through the collection of data relating to things such as actions items, decisions made, participation, time management, and attendance, meeting trends and patterns can be seen. This approach was adopted in the definition of the 'Task Progress' metric, described in Section 4.2.

4. Metrics developed

The metrics developed form two groups of metrics. The first concern the activities of 'idea generation' and 'enhancement or modification of existing ideas' and are described in Section 4.1. The second concern the activities of 'evaluation and filtering of ideas' and 'identifying and completing other types of design task' and are described in Section 4.2.

4.1. Metrics defined for effectiveness of idea generation

Based on the literature discussed in Section 3.1, it was clear that the metrics proposed by Shah et al (2003) are the *de facto* standard set of metrics for assessing the effectiveness of idea generation activities (quality, quantity, variety, novelty). However, it also became apparent that the metrics could not directly be applied to the range of co-design sessions considered within the SPARK project. In our research project we are not just looking at normal generative brainstorming sessions. We took the position that we do not know exactly what comes in out of co-design sessions due to the variance between the companies, projects and design stages. A system of metrics would therefore be needed that could cope with the variety of contexts that we might encounter. For example, the 'variety' metric proposed by Shah et al analyses concepts in a hierarchical manner using an 'Idea Genealogy Tree'. The tree is generated starting at the level of 'physical principles' proceeding down to 'working principles' then 'embodiment' and finally to the level of 'detail'. Concepts that share a common 'parent' at the level above are grouped together as 'siblings'. The Shah variety metric makes use of weightings to distinguish between a set of ideas that use the same physical principles with some variation at the embodiment or detail levels (giving a low score for variety) and a set of ideas that use several different working principles (giving a high score for variety). This approach is less applicable for the SPARK project as the platform will often be used during the embodiment and detail design phases, when the working principle has already been decided, thereby limiting the potential to differentiate idea sets using this variety metric.

The first stage in adapting the Shah metrics for use within the SPARK project was to utilise Morphological Charts (Zwicky, 1969) as the base framework for analysis. Morphological Charts were selected as they can help to identify changes in the 'solution space'. For the purpose of the SPARK project, Morphological Charts will not be used in the conventional way to generate ideas, but rather as a research tool to help analyse the creative design sessions. Before the creative design session being observed starts, the researchers will ask the facilitator of the session to explain each of the pre-defined concepts that will be presented to the non-designers in that session.

This information, along with the design representations used within the meeting, will be used by the research team to create a 'Before' Morphological Chart at the start of the session. Similarly, an 'After' Morphological Chart will be constructed by the research team for the sub solutions and concepts still under consideration at the end of the session. Building on the basic premise that morphological charts provide a sense of the size of the design space (Dym & Little, 2000), we propose that the difference between the 'Before' and 'After' Morphological Charts will show the evolution of the solution space that has occurred over the course of the co-design session. With the Morphological Chart providing the foundation for the analysis, we can now look at each of the metrics for the effectiveness of idea generation during a creative session.

4.1.1. Quantity of ideas metric

The quantity of ideas generated during the session is tracked by comparing the number of sub-solutions in the 'Before' and 'After' Morphological Charts.

Example:

Count of sub solution ideas in 'Before' Morphological Chart = 10

Count of sub solution ideas in 'After' Morphological Chart = 15

Quantity of new ideas = 15 – 10 = 5

4.1.2. Variety of ideas metric

For idea variety, the metric proposed is composed of two components: Variety (coverage) and Variety (new rows). Variety (coverage) is the proportion of the original feature rows in the 'Before' Morphological chart for which a new sub solution was generated. *Example*: Figure 2 shows a Morphological Chart in which sub-solution are represented by an 'X'. New sub-solutions generated during the session are represented by an 'N'. Number of features rows in 'Before' Morphological Chart = 4. Number of features rows for which a new sub solution idea was generated (Features A, C and D) = 3. Variety (coverage) score = 3/4 = 0.75

Note that a Variety (coverage) score of 1 would show that sub-solutions were generated for all of the original feature rows.

The Variety (new rows) component is defined as the number of new feature rows added as a proportion of the original number of feature rows. This is important in the scenario in which a sub-solution idea leads to the creation of a new feature row.

Figure 3 is similar to Figure 2 except that new sub-solutions were generated leading to the addition of two new feature rows (E and F) in the Morphological Chart. For this *example*:

Number of new feature rows added = 2, Number of original features rows = 4, therefore Variety (new rows) score = 2/4 = 0.5.

Features	**Sub-solutions**				
Feature A	X	X	X	N	
Feature B	X	X			
Feature C	X	X	X	N	N
Feature D	X	X	N	N	

Figure 2. Morphological chart example 1.

Features	**Sub-solutions**				
Feature A	X	X	X	N	
Feature B	X	X			
Feature C	X	X	X	N	N
Feature D	X	X	N	N	
(Feature E)	N				
(Feature F)	N	N	N		

Figure 3. Morphological chart example 2.

These two components, Variety (coverage) and Variety (new rows), are combined by simple summation to generate the overall variety score i.e. Variety = Variety (coverage) + Variety (new rows). *Example*: For the example shown in Figure 4, using the results for Variety (coverage) and Variety (new rows) calculated previously: Variety = 0.75 + 0.5 = 1.25

4.1.3. Quality of ideas metric

The approach to assessing idea quality proposed by Shah et al. requires each idea to be assessed against an established set of quality guidelines. These quality guidelines are generally specific to the type of product and type of project being assessed. Defining such quality guidelines may be feasible for design teams that regularly work with the same type of product, but it is not practical for design teams that encounter a large variety of different product types. Therefore, a simpler metric is proposed based on the number of new ideas generated that are taken forward at the end of the session for further development. This metric makes use of the fact that the participants of the co-design session will generally filter out low quality ideas and only take the higher quality ideas forward for further development. *Example*: Quantity of new ideas = 5, Number of new ideas rejected = 2, New ideas taken forward = 5 - 2 = 3

4.1.4. Novelty of ideas metric

The approach to assessing idea novelty proposed by Shah et al requires either: a database of possible solutions to compare with the solutions generated; or, the same task to have been completed by a number of different groups. Neither of these approaches is possible within the SPARK project context. An alternative approach, which has often been applied in the evaluation of creativity support tools, is to a panel of experts to judge the novelty of ideas. Therefore, the approach that will be applied in the SPARK project will be to take the mean average score for novelty from a panel of three experts for each of the ideas generated during the session. (Typically in design research, agreement between experts would have to be >70% and the coding scheme would need to be adjusted if this is not be achieved.)

Example:

Novelty score from expert A for idea 1 = 7	Novelty score from expert A for idea 2 = 5
Novelty score from expert B for idea 1 = 6	Novelty score from expert B for idea 2 = 5
Novelty score from expert C for idea 1 = 8	Novelty score from expert C for idea 2 = 6
Mean novelty score for idea 1 = 7	Mean novelty score for idea 2 = 5.33

Overal mean novelty score for session = (5.33 + 7)/2 = 6.17

4.2. Metrics developed for the idea evaluation, filtering and progress on tasks

Two further metrics have been defined to assess the effectiveness of co-design sessions – refered to as the ‘Filtering Effectiveness’ and the ‘Task Progress’ metrics.

4.2.2. Filtering Effectiveness metric

The ‘Filtering Effectiveness’ metric is a measure of the success of the idea filtering activities within a co-design session. Filtering of ideas is important to ensure progress in the project and therefore merits its own metric. The definition of the metric is as follows:

$$\text{Filtering Effectiveness} = \frac{\textit{Number of ideas rejected}}{\textit{Number of ideas considered} - \textit{Desired number of ideas to keep}}$$

Where ‘Desired number of ideas to keep’ is the number of ideas that should be left at the end of the session to take forward for further development. This number should be agreed by the facilitator of the session before the session starts or at the very start of the session. The ‘Number of ideas considered’

includes ideas brought in at the start of the session and any new ideas generated during the session. *Example:* Consider five ideas. Want to finish with one idea to take forward but the meeting only successfully filters out two ideas. Filtering Effectiveness = 2/(5-1)=0.5
Note that, a score of 1 would be a perfect filtering score.

4.2.1. Task Progress metric

This metric is based on the suggestion by Lucid Meetings (2015) to count the number of action items created and resolved. Here we use the term 'task' rather than 'action item', where a task is 'an activity the task doer performs in order to accomplish a goal' (Vakkari, 2003). The 'Task Progress' metric is calculated using a simple weighted score for each task resolved or created.
Task Progress = 3pts x (Number of high importance tasks resolved or created) + 2pts x (Number of medium importance tasks resolved or created) + 1pt x (Number of low importance tasks resolved or created). *Example*: For the example shown in Figure 4, Task Progress score = (3pts x 2) + (2pts x 3) + (1pt x 1) = 13 points.
The rating of tasks as being of 'high', 'medium' or 'low' importance is based on the strategic importance of the task to the success of the project. The rating will be conducted by the researchers initially (using standard best-pratice for meeting minutes note taking) and will be verified after the completion of the project by the team involved in the session. The weighting given to the task importance level was included because it was felt that a meeting which resolves or creates high-importance tasks is more effective than one that only deals with low-importance tasks. Resolving tasks and creating tasks are both considered to be positive in terms of meeting effectiveness (Lucid Meetings, 2015) as resolving shows that tasks are being completed, whilst creating tasks shows that more is now known about what needs to be done next, which is in itself progress.

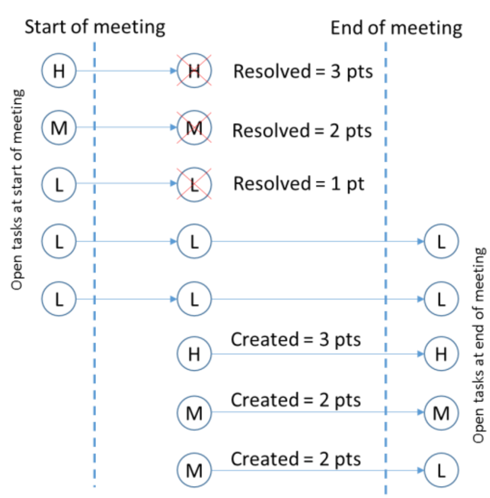

Figure 4. Example application of the Task Progress metric.

5. Conclusions

The aim of this paper was to propose a number of performance metrics that are suitable for co-design sessions. A list of the main types of activitiy that take place during co-design sessions was generated and used to guide a review of the academic literature, which identified a large number of metrics for the effectivenss of idea generation, but fewer metrics and less literature concerning other activities such as idea filtering or task progress.
The novel contribution of this paper is the definition of a broad set of metrics that can be used *together* to assess the performance and effectiveness of a range of different types of co-design sessions, from idea-generation focused sessions to review-focused sessions. Some of these metrics were based on established metrics but adapted for application within the SPARK project, whereas some are entirely new.
In the coming months, within the context of the SPARK project, the metrics defined here will be applied to a number of co-design sessions where design consultants were working with non-designers. Working with real companies on real projects has meant that the metrics have been developed to be applicable in those industrial settings whereas metrics proposed in literature for controlled studies can manipulate materials or the nature of the prototocol to be assessed. Subsequent papers from the SPARK project will report on the real-world testing of the metrics proposed, and the insights they offer.

Acknowledgement

The work reported in this paper was completed as part of the SPARK project, which has received from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417. This paper reflects only the authors' views and the European Commission is not responsible for any use that may be made of the information it contains.

References

Boujut, J.F. and Blanco, E., 2003. Intermediary objects as a means to foster co-operation in engineering design. Computer Supported Cooperative Work (CSCW), 12(2), pp.205-219.

Dym, C., and Little, P., 2000, Engineering Design: A Project-Based Introduction, John Wiley & Sons, Inc., New York.

European Union, (2014). Design for Innovation: Co-creation design as a new way of value creation, *Report by :Business Innovation Observatory, Case Study 14*, Ref. Ares(2015)4645055 - 28/10/2015

Füller, J., Mühlbacher, H., Matzler, K., & Jawecki, G. (2010). Consumer empowerment through internet-based co-creation. *Journal of Management Information Systems*, *26*(3), 71–102.

Geimer, J. L., Leach, D. J., DeSimone, J. A., Rogelberg, S. G., & Warr, P. B. (2012). Meetings at work: Perceived effectiveness and recommended improvements. *Journal of Business Research*.

Gosnell, C. A., & Miller, S. R. (2015). But Is It Creative? Delineating the Impact of Expertise and Concept Ratings on Creative Concept Selection. *Journal of Mechanical Design*, *138*(2).

Hannah, R., Joshi, S. and Summers, J.D., 2012. A user study of interpretability of engineering design representations. *Journal of Engineering Design*, 23(6), pp.443-468.

Leach, D. J., Rogelberg, S. G., Warr, P. B., & Burnfield, J. L. (2009). Perceived meeting effectiveness: The role of design characteristics. *Journal of Business and Psychology*, *24*(1), 65–76.

Lee, Y., 2008. Design participation tactics: the challenges and new roles for designers in the co-design process. *Co-Design*, 4(1), pp.31-50.

Lucid Meetings. (2015). Meetings Are Serious Business. Retrieved from www.lucidmeetings.com/resources

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220–232.

Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined metrics for measuring ideation effectiveness. *Design Studies*, *30*(6), 737–743.

O'Hern, M., & Rindfleisch, A. (2008). Customer Co-creation: A Typology and Research Agenda. *Review of Marketing Research*, (Working Paper No 4), 84–106.

Ostergaard, K.J., Wetmore, W. and Summers, J.D., (2003) A methodology for the study of the effects of communication method on design review effectiveness. In ASME 2003 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. 383-390).

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18.

Sanders, E. B.-N., & Stappers, P. J. (2014). Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign*, *10*(1), 5–14.

Shah, J. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111–134.

Shah, J. J., Kulkarni, S. V., & Vargas-Hernandez, N. (2000). Evaluation of Idea Generation Methods for Conceptual Design: Effectiveness Metrics and Design of Experiments. *Journal of Mechanical Design*, *122*(4), 377.

Taffe, S. (2015). The hybrid designer/end-user: Revealing paradoxes in co-design. *Design Studies*, *40*, 39–59.

Vakkari, P. (2003). Task-Based Information Searching. *Annual Review of Information Science and Technology*, *37*, 413–464.

Verhaegen, P. A., Peeters, J., Vandevenne, D., Dewulf, S., & Duflou, J. R. (2011). Effectiveness of the PAnDA ideation tool. In *Procedia Engineering* (Vol. 9, pp. 63–76).

Wetmore W.R., Summers J.D. and Greenstein, J.S. (2010), Experimental study of influence of group familiarity and information sharing on design review, effectiveness, *Journal of Engineering Design*, 21:1, 111-126.

Zwicky, F. (1969). *Discovery, invention, research: Through the morphological approach*. New York: Macmillan.