Measuring and comparing novelty for design solutions generated by young children through different design methods

Citation for published version (APA):

Sluis, R. J. W., Bekker, M. M., Eggen, J. H., Vermeeren, A., & de Ridder, H. (2016). Measuring and comparing novelty for design solutions generated by young children through different design methods. *Design Studies*, 43, 48-73. <https://doi.org/10.1016/j.destud.2016.01.001>

Document license:

TAVERNE

DOI:

[10.1016/j.destud.2016.01.001](https://doi.org/10.1016/j.destud.2016.01.001)

Document status and date:

Published: 01/03/2016

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Measuring and comparing novelty for design solutions generated by young children through different design methods

Wouter Sluis-Thiescheffer, Fontys University of Applied Sciences,
Rachelsmolen 1, 5612 MA Eindhoven, The Netherlands

Tilde Bekker and Berry Eggen, Eindhoven University of Technology, PO Box
513, 5600 MB Eindhoven, The Netherlands

Arnold Vermeeren and Huib De Ridder, Delft University of Technology,
P.O. Box 5, 2600 AA Delft, The Netherlands

When involving children in the design process, it is important to understand the novelty of their design solutions. This paper reviews the applicability of an often cited novelty metric Shah et al. (2003) for the comparison of two design methods conducted with 8–10 year old children. The novelty metric is applied to data that is different for a number of parameters, such as a different design phase (exploratory instead of conceptual), size and variety. The results yielded by this novelty metric are not straightforward. This paper describes the difficulties encountered and introduces an alternative approach. The alternative approach leads to better results for any amount of data, for an exploratory phase. Additionally the paper explains how this approach increases the sensitivity for detecting differences in novelty when comparing design methods.

© 2016 Elsevier Ltd. All rights reserved.

Keywords: design method(s), evaluation, novelty metric, reflective practice, user centred design

Corresponding author:
Wouter
Sluis-Thiescheffer
wouter.sluis@fontys.nl

Children are increasingly involved in the development of services and products. Since the beginning of this millennium, designing with children has become a recognized field, embodied in the annual ACM conference ‘Interaction Design and Children’ (since 2002). The field has reported many design methods that involve children in a design process (for example, [Druin, 1999](#); [Markopoulos, Read, MacFarlane, Hoysniemi, & Hysniemi, 2008](#)). The success of these methods has often been defined in terms of a successful process; reporting the parameters that keep children engaged and productive in generating ideas. Examples of such parameters are engagement and fun ([Bekker, Beusmans, Keyson, & Lloyd, 2003](#)), gender ([Hou, Komlodi, Lutters, Boot, & Cotton, 2006](#); [Stienstra, 2003](#)), group size ([Heary & Hennessy, 2002](#)) and power structures ([Heary & Hennessy, 2002](#);

www.elsevier.com/locate/destud
0142-694X *Design Studies* 43 (2016) 48–73
<http://dx.doi.org/10.1016/j.destud.2016.01.001>
© 2016 Elsevier Ltd. All rights reserved.

Pardo, Vetere, & Howard, 2005). The value of these methods in terms of the effectiveness of the ideas that children generate is often left unanswered. As the field of interaction design and children matures, it becomes relevant to evaluate the effectiveness of ideation methods for children. The purpose of involving children in the design process is, after all, to find innovative and relevant design directions for products that suit children in the best possible way.

Creativity is widely regarded as a key driver behind successful, innovative designs (Sarkar & Chakrabarti, 2011). The definition of creativity varies greatly and there is not one generally accepted definition. In the field of engineering design there is agreement over creativity as a problem solving skill (Galindo-Rueda & Millot, 2015; Hernandez, Shah, & Smith, 2010; López-Mesa, Mulet, Vidal, & Thompson, 2011). For a design solution to be regarded as a creative solution, novelty is considered a key aspect (Brown, 2014; Grace & Maher, 2014). Therefore, this paper focuses on novelty as a measure for creative value of design solutions. A method that inspires children to generate novel ideas is therefore considered more effective than a method that does not do that.

The proposed novelty metric in Shah, Vargaz-Hernandez, and Smith (2003) is part of a foundational and recognized work on ideation effectiveness. The work describes four metrics in total (novelty, variety, quality and quantity). The work of Shah et al. (2003) is widely adopted and has inspired researchers to propose further refinements and variations to the metrics, for example for the novelty metric (Hernandez et al., 2010; López-Mesa et al., 2011; Verhaegen, Peeters, Vandevenne, Dewulf, & Duflou, 2011; Wilson, Rosen, Nelson, & Yen, 2010) and the variety metric (Nelson, Wilson, Rosen, & Yen, 2009). These refinements consider additional levels of novelty (Verhaegen et al., 2011), the types of expectedness in novelty (novelty, surprise and transformational creativity, (Grace & Maher, 2014)) or add an aspect of review in the process of determining the novelty of an artifact (Sarkar & Chakrabarti, 2011). Brown (2014) explored the metric for application in the field of computational creativity. He found however that the metric is limited to Engineering Design, mainly because of elements included in the metric that rely on human assessment.

This paper describes the application of this novelty metric to compare the design solutions generated by children using two design methods, the nominal group technique (a form of brainstorming) and prototyping. The data was generated in a study reported in Sluis-Thiescheffer, Bekker, Eggen, Vermeeren, and De Ridder (2011). Interestingly, the metric proposed in Shah et al. (2003) only seems applicable to small datasets. Using the large dataset from Sluis-Thiescheffer et al. (2011) revealed two problems with the metric,

which made it very difficult, if not impossible to use this metric for that dataset. The first problem is that the (relative) metric only allows to compare design solutions within specific subsets (the attributes, see Section 2.3.2). The second problem is that applying the metric to a large dataset, decreases the difference in novelty between design solutions. A decreased difference in novelty between design solutions for large datasets is an important confounding factor in finding an effect. The cause of these problems probably lies in the modelling of novelty. A proper redesign of a function for novelty lies beyond the scope of this paper. Instead, we propose to use a simple procedure to compare design solutions and design methods on novelty, based on the top percentiles of least expected design solutions.

In previous work, we compared design methods based on the assessment of design spaces that result from the application of design methods with children. We have examined both a review- (Thang et al., 2008) and a metrics based approach (Sluis-Thiescheffer, Bekker, & Eggen, 2007) to assess the design spaces generated by children. A metrics based approach is preferred, because it is more efficient with large amounts of data, the method is less subjective and the results are inspectable (Vermeeren, 2009). This paper makes three contributions to the field. The first contribution is reviewing the calculation for novelty as proposed in Shah et al. (2003) (Section 2). The second contribution is introducing a different way to calculate the novelty metric (Section 3). The proposed metric starts from the same principle as Shah et al. (2003), that novelty is measured on the basis of frequency. The working principle of the metric is shown by using the dataset from a previously conducted comparative study (Sluis-Thiescheffer et al., 2011). The third contribution is applying the novelty metric in the field of participatory design with children, which has not been done before as far as we are aware (Section 5.5).

1 Study setup: a comparison between the nominal group technique and prototyping

In this section we describe the setup with which we compared the output of two early design methods.

When designing with young children, it is important to find a design activity that matches their skills and competencies (Sluis-Thiescheffer et al., 2007). An obvious example is that children need to have developed sufficient linguistic skills to participate in an interview. In Sluis-Thiescheffer et al. (2011) we introduced a framework that compares the skills acquired at a certain age with skills required for a design activity. Comparing required skills for a design method with acquired skills allows to hypothesize about the appropriateness of design activities for a certain age, given the acquired skills at that age.

Based on the framework we hypothesized that when a method involves the use of more, different skills, participants will be triggered in a larger variety of qualities and will have more room to express themselves, thus generating more ideas. When participants generate more ideas we expect to find more novel ideas, as we assume that the number of novel solutions among the trivial and obvious is a constant for a design method. Our study compares the output in ideas of a design method requiring relatively few skills with a design method requiring relative many skills. [Sluis-Thiescheffer et al. \(2011\)](#) lists a set of twenty-eight commonly used methods that require one to four skills.

According to the framework ([Sluis-Thiescheffer et al., 2011](#)) an NGT method (Nominal Group Technique, a variant of brainstorming – further explained in Section 1.1) would require mainly two design skills (linguistic skills and interpersonal skills) and a prototyping method (tinkering a low-fidelity but representative artefact) would appeal to four design skills (linguistic, interpersonal, bodily-kinaesthetic and visual-spatial skills). Since they are both commonly used design methods, it is interesting to explore the differences in ideation effectiveness with children. Using the hypothesis above, we expected that prototyping would result in more attributes and in more creative ideas than NGT.

The study reported in [Sluis-Thiescheffer et al. \(2011\)](#) (examining differences in quantity), showed that with the prototyping method, the children indeed generated significantly more ideas than with NGT. Interestingly, the ideas generated with NGT were, according to expert assessment, more creative [Thang et al. \(2008\)](#). One explanation could be that the ideas generated through NGT were more novel than the prototyping method. Since novelty is an important factor in an exploratory design phase (the design phase we focus on) we want to examine this possibility. To be more efficient than a labour-intensive extensive expert review method, we will examine the data by using a metric for novelty. The metric is used to determine whether or not the ideas generated through NGT are indeed more novel than through prototyping. Before going into the application of the novelty metric, we will introduce the study setup.

1.1 Nominal group technique as a brainstorming method and a prototyping method

The form of brainstorming we used is The Nominal Group Technique (NGT) since it resembles a widely used brainstorm technique. A brainstorm exists in many variations (e.g. [Michalko \(2006\)](#)). [Osborn \(1942\)](#) defines a brainstorm as.

“a conference technique by which a group attempts to find a solution for a specific problem by amassing all the ideas spontaneously generated by its members.”

However, it is also known that when brainstorming in a group, the stream of ideas quickly dries as a consequence of cognitive tuning (Fern, 2001). Cognitive tuning is the phenomenon that members of a group ‘tune into’ each others mindset, limiting the diversity in ideas. The NGT is therefore a good alternative. The NGT allows the group members to develop their own ideas, before sharing them with a group (Fern, 2001) hence avoiding cognitive tuning.

Prototyping is defined by the definition of Muller (2003):

“Creating an understanding for an artefact and changing the artefact to explore understandings of one another’s positions, to question one another’s approaches and to accommodate heterogeneity of views and interests.”

1.2 Design problem: attending class from a remote location

A telecommunication company provided a design problem they were working on, a device for children in a primary school. The goal of the design case was to establish live communication between a child at home and the class in school. The purpose of the live connection was to facilitate attending class for a child that has almost recovered from an illness. The child is not yet able to come to school, but is mentally fit enough to attend at least some hours of class. We want the children to generate ideas about a device located in class and one at home. The design space was focused on finding interface solutions in the home situation and at school. The school situation was further specified for two common class situations, thus providing the two design cases. In one case the focus was on a collaborative activity, where the child would work in a small group of peers. In the other case, the focus was on a class activity where the child pays attention to the teacher. The teacher is actively teaching in front of the classroom.

1.3 Focus group setting

As explained above, the basis for comparing NGT-output with prototyping output is the conversations about the objects and actions of the device. For the set-up of the study we have to choose between group conversations and one-on-one (child-experimenter) conversations. Hennessy and Heary (2004) report that in a focus group the dynamic interaction of five children will elicit more information than a one-on-one interview because there is less emphasis on the adult-child relationship. Moreover, design activities with user-participants are often reported to be group-activities (Langford & McDonagh, 2003). For the group size, Hennessy and Heary (2004) report that a focus group session is performed best with a maximum of five children. A group of five children is large enough to have at least

one talkative child, and yet small enough to give each child sufficient attention.

Because of the structure of a focus group conversation, the results can be analysed on a group level and on a per child basis. On the group level, all observed options and criteria during the group conversation are taken into account. However, although the group-design has its advantages for the involvement of children, the group work could average out the differences that are the focus of this study. To avoid that, both a group analysis and an analysis on the basis of each child's individual explanation are required.

1.4 Participants

Twelve groups of ten-year-old children participated, ten groups of five participants and two groups of 6 participants (a total of sixty-two children). All groups were of a mixed gender composition (at least two of each gender). All children were in grade five of a Dutch primary school. Three primary schools, one from the centre of the Netherlands and two schools from the south of the Netherlands took part in the study.

1.5 Within-subject design

A within-subject design was applied. Each group of five children participated in both an NGT-session and a prototyping session. The study was designed to avoid potential carry-over effects of either session by offering the NGT-session and the prototyping session in alternating order between the groups. A typical session would have the following setup:

1. General introduction
2. Introduction of the first case
3. Warming up for idea generation
4. Individual time to write ideas (NGT)
5. Group discussion on the written ideas
6. Introduction of the second case
7. Warming up for idea generation
8. Individual time to create a prototype
9. Group discussion on the prototypes
10. Rounding off

To treat talkative and less talkative children alike and to avoid a bias in (accidentally) focussing the discussions on specific ideas or prototypes by the experimenter, the experimenter was allowed to only ask two questions per child. Each method dealt with a different design case to avoid a feeling of repetition among the participants. The design case proved not to have an effect on the quantity of the design output (Sluis-Thiescheffer et al., 2011).

2 Review of applying the novelty metric S

To measure the novelty of ideas in a design space, we will first define what we mean with an idea, then define novelty and finally introduce the process of applying the metric for novelty as proposed in [Shah et al. \(2003\)](#).

2.1 Defining novelty

2.1.1 Definition of a design idea as a design artefact composed of attributes

[Shah et al. \(2003\)](#) take an approach on measuring novelty with a design method that results in design artefacts. Thus, the ideas resulting from the method are in fact the artefacts. They provide an example of a prototyping session which results in physical objects that address the design problem. The results of a brainstorming session, as used in this paper, are not artefacts per sé, but design artefacts in verbal descriptions of physical objects and actions. To include this method and other methods that result in a description of a design artefact, like sketching sessions, or roleplaying, we broaden the term design artefact to include verbal descriptions of physical objects and actions.

[Shah et al. \(2003\)](#) explain that a design artefact is further specified by its attributes and functions. In this paper we will continue to use that terminology with one simplification: the term ‘attribute’ will be used to refer to both attributes and functions. Although there are semantic differences (e.g. properties vs actions), both functions and attributes are treated identically in the calculation of novelty.

Finally we distinguish between an attribute as a category and as an instance of that category. The term ‘*attribute*’ in the remainder of this document, will refer to the *category*, the term ‘**design solution**’ will refer to an instance of that attribute. For example, in [Shah et al. \(2003\)](#) the design method resulted in design artefacts that could travel as far as possible, propelled by compressed air. Each device had an attribute ‘*propulsion*’. Some devices were propelled by ‘*sail*’, others by ‘*jet*’ etc. The instances ‘*sail*’ and ‘*jet*’ are called ‘**design solutions**’ for the attribute *propulsion*.

To summarize the breakdown of the content of a design method:

- A design method results in design artefacts
- Each design artefact is described with attributes
- Each attribute is described by the design solutions (instances from the corresponding design artefacts).

2.1.2 Definition of novelty

Novelty is defined by [Shah et al. \(2003\)](#) as:

a measure of how unusual or unexpected an idea is, as compared to other ideas.

In other words, novelty is defined as a measure of expectedness, where unexpected ideas are considered novel and expected ideas are not-novel. The novelty value (i.e. the expectedness) of a design solution is determined a posteriori: it is calculated based on the frequency of that design solution within the generated design space. More specifically: the novelty score is determined by the number of times a specific design solution is generated compared to the total number of design solutions generated for an attribute. The novelty score for the design artefact is the sum of the novelty scores of the corresponding design solutions. The novelty score for a method is the sum of the novelty scores of the design artefacts in the generated design space.

2.1.3 Five steps to compare design methods in terms of novelty

The general process for applying the novelty metric in [Shah et al. \(2003\)](#) consists of five steps:

1. collect all design artefacts from all methods;
2. decompose the design artefacts into instances of the main attributes and functions (i.e. into attributes);
3. apply the novelty metric to determine the novelty score for the attribute;
4. determine the novelty score for the design artefact by a summation of the novelty scores for attributes (i.e. the instances of attributes and functions);
5. determine the novelty score for a method by a summation of the novelty scores for design artefacts.

2.2 The novelty calculation according to Shah et al.

The novelty value (S) for each design solution is calculated according to Equation (1). T_j is the total number of design solutions from all participants from all methods for attribute j ; C_{1j} is the frequency of design solution one for attribute j and S_{1j} is the resulting novelty value for design solution one for attribute j . According to this equation, the resulting S can be any number between zero and ten.

$$S_{1j} = \frac{T_j - C_{1j}}{T_j} \times 10 \quad (1)$$

2.2.1 Example from Shah et al.

Shah et al. applied their novelty metric to 46 design artefacts from a design competition. The design competition asked for an object that could travel as far as possible, somehow propelled by compressed air. For the design artefacts

they identified four attributes. Each design artefact satisfied each attribute with one solution, hence Shah et al. worked with four times forty-six attributes. The four attributes were (1) *Propulsion*, (2) *Medium of Travel*, (3) *Motion* and (4) *Number of Parts*.

An overview of their dataset is presented in Table 1. The total number of design solutions for an attribute is referred to with ‘T’. The number of different design solutions is ‘t’, for example: attribute propulsion has the design solutions ‘jet’, ‘jet on sail’, ‘turbine’ etc. The *S* interval shows the novelty values for the least and most frequently generated solutions for an attribute.

In this dataset there were only two design artefacts that had ‘Water’ as a design solution for the attribute ‘*Medium of Travel*’. *Water* is therefore the most novel solution ($C = C_{min} = 2$). The novelty value for water is calculated in Equation (2).

$$T_{MediumofTravel} = 46, C_{water} = 2 \Rightarrow S_{water} = (46 - 2) * 10 / 46 = 9.56 \quad (2)$$

2.3 Review of applying the novelty metric *S* to compare methods

First we will describe a number of problems we encountered during the process of applying the method to the data from the comparative study. Then we will explain these phenomena by means of two theoretical, extreme examples.

2.3.1 Problems detected when applying the novelty metric to the dataset from the comparative study

As mentioned in the introduction of this article, applying the novelty metric by Shah et al. (2003) resulted in novelty values that were difficult to interpret. In this section we will show a summary of the factors under which the datasets were obtained and run through a visual inspection of the data when we applied the novelty metric as introduced in Shah et al. (2003). With the visual inspection of the data we can point to the problems found to justify a review of the novelty metric. In subsequent sections will propose a different calculation that overcomes these problems.

In our study we focused on an early phase in a design process, an exploratory phase. We collected verbal explanations generated by children of around ten years of age through two defined ideation methods: prototyping and brainstorming. In Table 2 we present an overview of the characteristics of the data collected and compare them to the characteristics of the design competition in Shah et al. (2003).

There are differences in settings, like the age of the participants and the design phase. Although the means of expression are different, the attributes and

Table 1 Overview of the dataset presented in [Shah et al. \(2003\)](#). *T* is the total number of design solutions for that attribute, *t* is the number of different design solutions, *C max* is the frequency of the most often (most trivial, least novel) generated design solution, *C min* is the frequency of the least often (least trivial, most novel) generated design solution for an attribute. The *S interval* shows the interval of the novelty scores for an attribute

Dataset Shah et al. (2003)

Attribute		T	t	C max	C min	S interval
1	Propulsion	46	10	26	1	4.3–9.8
2	Medium of travel	46	3	34	2	2.6–9.6
3	Motion	46	3	30	3	3.5–9.3
4	Number of parts	46	2	39	7	1.5–8.5

Table 2 The characteristics of the study setup in [Shah et al. \(2003\)](#) and of the comparative study

Study setup	<i>Shah et al. (2003)</i>	<i>Comparative study</i>
Number of participants	46 Teams	62 Children
Type of participants	Design students	Children of around 10 years old
Design phase	Concepting	Exploratory / Idea generation
Number of methods	1 (Prototyping)	2 (Prototyping and brainstorming)
Medium of expression	Prototypes	Verbal explanations; supported by prototypes or brainstorming notes

artefacts are similarly articulated. In [Shah et al. \(2003\)](#) the analysts interpreted and explicated the attributes, in our study, the children themselves explained what the attributes of their ideas were. Both design studies resulted in data that is suitable to apply the novelty metric to: 46 artefacts (one per team) in [Shah et al. \(2003\)](#) and 62 design ideas (one for each child) in [Sluis-Thiescheffer et al. \(2011\)](#).

The ideas in the comparative data study were verbally explained design artefacts. The design artefacts explained in the dataset were analysed according to the proposed procedure in [Shah et al. \(2003\)](#). To compare the novelty values all design artefacts, generated by all participants from all methods were collected. Then we identified the attributes that were relevant for solving the design problem, such as input device, output device etc. The total number of solutions for an attribute is referred to with ‘T’. Then we identified the unique solutions for those attributes, i.e. ‘t’, the number of different attributes (for example: input device = keyboard, microphone, camera). Finally, the frequency of each unique solution, determined the novelty of each solution for an attribute. [Table 3](#) provides an overview of the numbers. The *S*-interval shows the novelty values for the least and most frequent solutions for an attribute.

Visual inspection of the *S*-intervals shows all solutions generated have a rather high novelty value: none of the intervals drops below the 6.3. This can mean different things, among others that the children were highly creative. It could

Table 3 The dataset resulting from our study. *T* is the total number of design solutions for that attribute, *t* is the number of different design solutions, *C max* is the frequency of the most often (most trivial, least novel) generated design solution, *C min* is the frequency of the least often (least trivial, most novel) generated design solution for an attribute. The *S interval* shows the interval of the novelty scores for an attribute

<i>Dataset comparative study</i>						
<i>Attribute</i>		<i>T</i>	<i>t</i>	<i>C max</i>	<i>C min</i>	<i>S interval</i>
1	Human device interaction	816	170	160	1	8.0–9.9
2	Information	505	124	64	1	9.1–10
3	Physical property	374	49	137	1	6.3–10
4	User	350	19	131	1	6.3–10
5	Location	293	58	53	1	8.2–10
6	Input device	265	43	74	1	7.2–10
7	Output device	217	40	80	1	6.3–10
8	Throughput device	199	44	41	1	7.9–9.9
9	Non-device object	193	102	19	1	9.0–9.9
10	Time	13	6	4	1	6.9–9.2

also mean that something particular is going on with the novelty calculation. One factor is the number of different solutions (*t*), the higher the number of different solutions, the smaller the *S*-interval and closer to perfect novelty $S \sim 10$. It is hard to conceive though that the least novel idea for example for Human Device Interaction, still scores an 8.8. There are 816 solutions (Table 3), and the raw results showed that the least novel contribution has a frequency of 160. Thus a solution that has been generated on average 2–3 times by each of the 62 subjects receives a higher novelty score than for example the least novel solution for time. The least novel solution for time (generated 4 times, received a novelty score of 6.9), was generated on average by only 1 in every 16 participants. The phenomenon is not limited to our data, it is also the case for the novelty values in Shah et al. (2003) in Table 1 albeit less strong, because they have a lesser amount of data.

The next section will explain this phenomenon further by means of extreme examples. After that we will continue to explain an alternative approach that resolves this problem.

2.3.2 The problems with *S*-scores explained with two extreme examples

The high novelty scores for trivial solutions raises the question of how the novelty values should be interpreted. In Shah et al. (2003) the authors do not explicitly explain how the *S*-values should be interpreted. In the description of the working principle, Shah et al. assume that all *S*-values from all attributes are equally meaningful when they add the novelty values of solutions for different attributes to calculate the overall novelty of a design artefact. Their assumption implies a model of novelty on a 10-point, continuous scale, because of the reach of $S \in [0 - 10]$ where $S = 0$ means completely trivial

and $S \sim 10$ approximates ultimate novelty. The next paragraphs show the dynamic behind the high novelty scores for trivial solutions. Based on that we will argue that the distribution of S values for solutions of an attribute needs to be considered before S -values for solutions of different attributes can be added up. This can be best illustrated by means of extreme examples.

Suppose there are forty design artefacts for which we examine two attributes (P and Q). Solutions for the attributes are generated through two methods (A and B). Method A resulted for attribute P in forty times the same solution X ($T_P = 40$ and $C_{PX} = 40$). The novelty value for solution X is $S_{\text{solution X}} = (40 - 40)/40 \times 10 = 0$ (see Figure 1). Considering the theoretical reach of S , $S \in [0 - 10]$ this result indicates that method A was extremely unsuccessful: the one solution generated receives the lowest possible novelty score of $S_{PX} = 0$.

Method B resulted for attribute P in forty unique solutions ($T_P = 40$, $C_{\text{any solution}} = 1$). The novelty value for any of the solutions is the same, because they all have a frequency of $C_{\text{any solution}} = 1$. Suppose one of the solutions of method B coincides with method A. The novelty score for solution X using method B is $S_{PX} = (40 - 1)/40 \times 10 = 9.8$ (see Figure 2). Considering the reach of S , $S \in [0 - 10]$, this result indicates that method B was extremely successful: the resulting solutions approximate a near perfect novelty score of $S_{P \text{ any solution}} = 9.8$.

Thus, the same solution X receives either the maximum or the minimum novelty value, depending on whether it was generated through method A ($S_{\text{attribute XA}} = 0$) or method B ($S_{\text{attribute XB}} = 9.8$). In the process of comparing methods, the set of solutions for attribute P from method A and B are merged into one set of solutions. The merging process changes the novelty score of solution X. In the merged set of solutions for attribute P ($T_P = 80$), solution X ($C_{PX} = 41$) has a novelty value of $S_{PX} = (80 - 41)/80 \times 10 = 4.9$. Figure 3 shows how the changed distribution of solutions in the combined set changes the novelty values accordingly.

This extreme example suggests that the S -values can only be interpreted within the context of the distribution of novelty values in the set: $S_{PX \text{ combined}} = 4.9$ and $S_{PX \text{ method A}} = 0$ are different values, that share the same meaning: ‘solution X is the most trivial solution’. The value of S_{PX} therefore cannot be interpreted without knowing the distribution of novelty values for attribute P in either set.

Interpretation of the novelty values becomes more complicated when a second attribute (Q) is taken into consideration. Suppose method A resulted for attribute Q in two different solutions, Y and Z, both solutions were generated twenty times. The novelty score for solution Y using method A is

Extreme Example 1
Method A results in one unique design solution (C=40, S=0)
for attribute P (T=40)

Figure 1 An extreme example of design solutions generated through method A for attribute P. With each method, 40 design solutions were generated. Method A generated 40 times the same solution

$S_{QY} = (40 - 20)/40 \times 10 = 5$. Because solution Y and Z have the same frequency, the novelty score for solution Z is also twenty; $S_{QZ} = (40 - 20)/40 \times 10 = 5$. Considering the reach of $S(S \in [0 - 10])$ indicates that the solutions generated with method A are mediocre, a novelty score of five approximates the middle between the extreme ends of that reach. The distribution of frequencies is not informative in this case. All ideas are equally novel, and there is not a relative minimum or maximum novelty value. Therefore it is unclear whether Y and Z are the two most novel or the two most trivial solutions.

Suppose method B yielded the same results; identical to method A, method B resulted in two solutions for attribute Q, Y and Z, each was generated 20 times. Thus method B also yielded $S_{QY, QZ} = 5$. To compare the two methods for attribute Q, the two datasets are merged. Solution Y and Z in the combined set maintain the same novelty value as in set A or B of $S_{QY, QZ} = (80 - 40)/80 \times 10 = 5$, indicating a mediocre result based on the reach of S. Again, the distribution of the values does not provide more information, as all novelty values are the same.

The novelty score for attribute Q illustrates raises another question: How does the novelty value of the solutions Y and Z ($S_{QY, QZ} = 5$) relate to the novelty value of solution X ($S_{PX} = 4.9$)? If solution X should be interpreted as the least novel solution, and solution Y as a mediocre solution, then the novelty values for the solutions for attribute P require another operation to reflect that difference. If solutions X, Y and Z should be interpreted as more or

Extreme Example 2
Method B results in 40 unique design solutions ($C=1, S=9,8$)
for attribute P ($T=40$)

Figure 2 An extreme example of design solutions generated through method B for attribute P. With each method, 40 design solutions were generated. Method B generated 40 times a unique solution

Extreme Example 1 and 2
The design solutions for attribute P ($T=80$) of method A and B are combined;
the novelty value for the design solutions of method A is now $S=4.9$.

Figure 3 When the set of solutions generated through method A is merged with method B, the solutions for attribute P in method A increase in novelty from $S = 0$ to $S = 4.9$. For attribute P, the lowest novelty score in the combined set is now 4.9

less equally trivial results, the difference between the original values for solution X (Method A: $S_{PX} = 0$ vs $S_{QY} = 5$) are hard to explain. So, within each attribute and within a method, the interpretation of novelty scores makes sense. A comparison between methods using S is hard, if not impossible. The changes in novelty scores between the combined set for method A and B and the original sets for method A and B respectively will make it hard, if not impossible to trace coherent and consistent differences and subsequently interpret and explain those differences between the methods.

3 *The proposed novelty calculation*

To overcome the indicated problems, this paper proposes a simple but effective method to compare methods for the novelty of design solutions. The proposed procedure is based on a binary approach for determining the novelty (B) for a solution, a solution is either novel or not novel. The criterion for being novel is determined by meeting a frequency threshold or threshold of expectedness. A solution that is generated more often than the à priori defined expectedness threshold is not novel, and receives a novelty score of $B_{i,j} = 0$. Consequently, a solution that is generated less frequent than the threshold is considered novel, and receives a novelty score of $B_{i,j} = 1$. A threshold in terms of percentiles of the total number of generated solutions takes into account the size of the dataset.

3.1 *Example from the current study*

To determine whether or not an attribute is novel, we set the frequency threshold à priori at a .75 percentile (Q3) for the frequencies of the solutions for an attribute (i.e. Q3 of all values for C_j in Equation (1)). The choice for Q3 is in principle arbitrary, but it seems logical to choose a value that disregards at least 50% of the most frequently generated solutions. The third percentile focuses on the twenty-five percent of least frequent half of generated solutions. To assign the novelty value to the solutions, we first rank the solutions by frequency, from the highest frequency to the lowest frequency, i.e. from the least novel solution to the most novel solution. Then we determine Q3. A solution with a frequency higher than Q3 is an expected solution, i.e. it is a non-novel solution with $B_{i,j} = 0$. A solution with a frequency equal to or lower than the Q3 frequency is an unexpected solution, i.e. it is a novel solution with $B_{i,j} = 1$.

Figure 4 shows the attributes for the attributes Time and HDI in order of descending frequency. The figure shows that for Q3 the novelty threshold $Q3_{Time} = 2$, hence the attributes with a frequency of two and lower ($C_j \leq 2$) are considered novel. For HDI the novelty threshold $Q3_{HDI} = 4$ (hard to read in the figure because of the many data points), hence the attributes for HDI are considered novel with a frequency of four and lower ($C_j \leq 4$).

1. Collect all design artefacts generated;
2. Identify for each design artefact how the attributes are satisfied (i.e. observe the solutions);
3. Set the ‘expectation threshold’
4. Calculate for each solution h (in attribute i , from design artefact j) whether $B_{h,i,j} = 1$ or $B_{h,i,j} = 0$.

1. Calculate the sum of the novelty scores for each design artefact ($M_j = \sum B_{h,i}$)
2. Compare the scores for each design artefact to determine the most novel design artefact.

63

1. Calculate the sum of the attributes ($\sum B_{i,j}$) per session for each method
2. Compare the scores for each session for each method.

4 Results of comparing NGT and brainstorming for novelty

This section reports the results for comparing the two methods for novelty.

4.1 No significant difference between the number of novel solutions for the two methods

The twelve group discussions were analysed for ten different attributes (see [Table 3](#)), using the steps explained in the previous section. In total there were 3225 solutions generated through the combination of prototyping and NGT. After allowing only the 25% least frequent solutions for each attribute, conversations after the NGT session contained on average 40 novel solutions with a .05 confidence interval of [31.419; 48.248]. The conversations after the prototyping session contained on average 38 novel solutions with a confidence interval of [27.947; 47.053]. Using repeated measures did not show a significant difference between the two methods. [Figure 5](#) shows the difference between the average number of novel solutions.

4.2 Significantly more novel solutions when prototyping follows NGT

The group results were further analysed for possible interaction effects. Using repeated measures with $\alpha = 0,05$ shows a between subject effect for order of $F = 5,926$ and $p \leq 0,035$. [Figure 6](#) shows the difference in average number of novel solutions for order.

5 Discussion

In this paper we applied the novelty metric as proposed in [Shah et al. \(2003\)](#) to field data. The field data were collected from a study into design solutions generated by children with two different methods: NGT and prototyping. In the process of applying the novelty metric, we found that a consistent interpretation of the novelty scores was difficult using the proposed metric and we presented an alternative process that does allow a consistent interpretation. Furthermore we have shown that our approach simplifies the procedure of applying the novelty metric. The resulting novelty scores for the study did not result in a significant difference in the number of novel attributes between NGT and prototyping. The study did show a significant difference in the number of novel attributes depending on the order in which NGT and prototyping are applied.

Figure 5 The average number of novel solutions generated per method in a design session, either after an NGT session or after a prototyping session. The error bars in the figure show the .05 confidence interval, using repeated measures of the general linear model. To display a representative confidence interval for a within-subject contrast, the confidence interval is divided by $\sqrt{2}$ (Lof-tus & Masson, 1994)

Figure 6 The average number of novel solutions generated per method in a design session, either after an NGT session or after a prototyping session. The figure shows a between subject effect for order; the order NGT-prototyping results in more novel solutions than the reverse order. The error bars in the figure show the .05 confidence interval

5.1 Application of the metric on field data revealed an interpretation problem with the novelty metric

In the demonstration of the novelty metric in Shah et al. (2003) the authors did not encounter the reported problems. There are at least two good reasons: (1) they did not have to show the procedure for comparing methods to show the

working principle for the novelty metric and (2) they worked with entries from a conceptual design phase resulting in an identical number of design solutions for each attribute, whereas we worked with data from an exploratory design phase, resulting in different numbers of design solutions for each attribute.

The proposed metric for novelty by [Shah et al. \(2003\)](#) is fundamental in nature. A demonstration of interpreting the results of the novelty metric beyond comparing design artefacts is not required for demonstrating a working principle. There are others that considered ([López-Mesa et al., 2011](#)) or applied the metric in a study ([Ferent & Doboli, 2011](#); [Hernandez et al., 2010](#); [Wilson et al., 2010](#)), but none that we are aware of applied the metric for comparing design methods.

Furthermore, there are two differences between the [Shah et al. \(2003\)](#) dataset and our dataset that can explain why we detected the identified problems (see [Table 2](#) in [Section 2.3.1](#) on page 2: (1) they used a low number of design artefacts and (2) our design artefacts were generated in a more exploratory phase of the design process whereas in [Shah et al. \(2003\)](#) their design artefacts were created during a conceptual phase. Shah et al. analysed 46 design artefacts for which they identified four attributes. Each artefact satisfied each attribute with one attribute, hence Shah et al. worked with 4 times 46 attributes.

As explained in [Section 2.1](#) one difference is that in [Shah et al. \(2003\)](#) each design artefact had the same number of relevant attributes and for each attribute there were 46 design solutions. This is a situation that is likely to happen in a design competition where a design artefact has to comply with certain rules. Furthermore it is a probable situation in the context of conceptual design phase, in which the important attributes for a design are determined. In an exploratory design phase, in which solutions are explored, the number of attributes and the importance of each attribute is not (yet) determined. In the design sessions, some children focused on a few specific attributes, generating multiple design solutions for these specific attributes and neglecting the other attributes. As a result, the design artefacts from the presented study are not equal in the number of solutions for each attribute. The exploratory character of the design artefacts is explained in [Figure 7](#).

Normalization is briefly addressed in [Shah et al. \(2003\)](#), they mention the multiplication by 10 in the formula for S as a means to normalize. However it is unclear in what way the novelty metric is normalized. Considering the formula for novelty S , it probably means that the data is normalized to fit a 10 point scale. In their example it looked like all the novelty values were comparable and interpretable on a 10 point interval. However, in our review, we found that the novelty values of different attributes are in fact not comparable. An identical number of solutions for each attribute, and a relatively low number of data points can obscure that. A varying number of solutions and a large

(a) Multiple solutions for one attribute

(b) Missing an attribute

Figure 7 In contrast with the dataset from [Shah et al. \(2003\)](#), in the results of the comparative study, the number of design solutions per attribute differs. Some children suggested alternative design solutions for one attribute ([Figure 7\(a\)](#)). Some children included design solutions for attributes (e.g. location of the design artefact) that others disregarded ([Figure 7\(b\)](#))

dataset revealed that the metric would require a normalization of scores between attributes, at least when comparing novelty scores from multiple attributes in artefact or method comparisons.

The proposed method works around the issue of creating normalized novelty scores for the entire solutions space. The proposed method only focuses on the most novel ideas, which implies that in this area of the design space the differences in novelty between attributes for each attribute are small enough to be neglected. The congression of the lines for each function around $N = 10$ in [Figure 8](#) shows that in the [Shah et al. \(2003\)](#) study the differences are at least smaller when working with the most novel attributes. The chosen threshold value (in our approach the 25% most novel ideas) determines what is considered an acceptable level of relative difference in novelty between the attributes to work with. Some kind of normalization to represent the difference in frequencies would have been a more elegant solution. Normalization introduces the question of what normalization function should be applied, which subsequently gives rise to a more fundamental question; beyond the scope of this paper: How does novelty relate to frequency?

Figure 8 Novelty as proposed by [Shah et al. \(2003\)](#) can be represented as a linear function of the frequency of attributes. The figure shows the function for the attributes in [Shah et al. \(2003\)](#) and for the attributes time ($n = 13$) and human–device interaction ($n = 816$) from our dataset

5.2 A significant interaction effect for NGT followed by prototyping

The results from the experiment show that the average number of novel design solutions generated through NGT does not differ significantly from the average number of novel ideas generated through prototyping. A comparison of the averages indicates that prototyping generated a lower number of novel design solutions than the NGT method. This indication is consistent with what we found through peer review ([Thang et al., 2008](#)). In that study the same dataset was used and the described design artefacts were reviewed by design experts. The experts considered design artefacts generated through prototyping consistently less novel than design artefacts generated through NGT. The data points in that study were large enough for a straight comparison, but not large enough for measuring a significant interaction effect for the order of design methods.

Interestingly, this study shows that there is a significant interaction effect for order. When NGT was performed first, there is a significant increase in novel design solutions generated through prototyping, whereas when prototyping was performed first, the subsequent NGT method did not result in significantly more novel design solutions. Only a few studies that we are aware of investigated the effects of brainstorming on subsequent creative tasks ([Al-khatib, 2012](#); [Coskun & Yilmaz, 2009](#); [McGlynn, McGurk, Effland, Jöhl, &](#)

Harding, 2004). The general finding is that brainstorming has a positive effect on subsequent idea generation, or creative problem solving tasks. McGlynn et al. (2004) also shows that the earlier evidence for possible design directions is shown in a series of subsequent brainstorming tasks, the more the performance in subsequent brainstorming task becomes inhibited.

The effect might be explained from the theory of multiple intelligences. In our framework (Sluis-Thiescheffer et al., 2011) we hypothesize that the more intelligences are involved in a design method, the more perspectives on a design problem are triggered. The more perspectives on a design problem are available, we expect that the possible combinations will lead to the generation of more creative solutions. This hypothesis was rejected in Thang et al. (2008), because experts considered ideas generated through prototyping significantly less creative. The explanation is that the multi-dimensional design space, triggered by a higher number of involved intelligences leads to triangulation. The process of triangulation evaluates an idea among multiple perspectives, inhibiting ideas that seem impossible from one of the perspectives. The interaction effect however, leads to the idea that brainstorming makes the more novel solutions generated through prototyping more focused. A focus on a specific design space allows an in depth exploration of that design space, thus limiting the inhibition by triangulation. Further research is required to better understand the relation between brainstorming and subsequent ideation techniques.

5.3 The value of the threshold

In our proposal we suggested a threshold at 25% of the number of the most novel solutions. That seems a sensible approach when working with large amounts of data. However, when there are only a few solutions for an attribute (like for Time in our dataset), the threshold might seem somewhat arbitrary as the solutions differ only one instance in frequency (see Figure 4. For smaller datasets, it is probably better to choose the median (50% of the number of solutions with the lowest frequency). Further research is required to develop proper guidelines on which threshold to apply. An other solution is to create a more weighted value that accounts for the differences in frequency within the percentile.

The advantage of disregarding 75% of the data is that it resolves the problems encountered. The bias of high novelty scores for trivial ideas will not apply. In the extreme case for attribute P where method A resulted in trivial design solutions and method B resulted in only novel design solutions, all design solutions for method B except design solution X will be included in the novelty score. In a method comparison, B results in 39 novel design solutions and A results in 0 novel design solutions, clearly showing B to be the most effective for generating novel (unexpected) ideas. In the extreme case for attribute Q, where both method A and B resulted in the same number of design solutions

Y and Z, it will depend on the threshold whether the design solutions are considered novel or not. The design solutions Y and Z each accounting for half of the generated ideas, will only be considered novel if the threshold for novelty is set to 50%. A higher threshold will disregard the ideas as being too trivial to be considered novel (which makes sense if they are generated so often).

Another disadvantage of using a threshold of 25% is that 75% of the data is disregarded. There is room for improvement there, but it requires further advancements in computational novelty. In the current situation, the long tail of common, or even trivial solutions only diminishes effect measures on the low number of novel solutions, and therefore we suggest to focus only on the most novel solutions in the data.

5.4 Expectedness as a measure for novelty

Expectedness is a simple, yet straightforward representation for novelty of design solutions (Grace & Maher, 2014). The results show that the metric for computed novelty did not yield a significant difference between the attributes generated through NGT and prototyping, whereas designers ratings (Thang et al., 2008) do differentiate the methods on assessed novelty. The lack of a difference could be explained by the basis for the computed novelty metric: expectedness. Expected solutions are probably indeed not-novel as they occur in a higher frequency. Unexpected solutions on the other hand are probably not only novel solutions, but possibly also trivial ones. Trivial solutions are rarely expressed, because they are well known, existing solutions. Good conversation skills require to minimize trivial contributions, which also results in a low frequency. The low frequency results in a high novelty score for trivial solutions. Designers can distinguish between novel and trivial solutions, however, in applying the metric there is no such filtering. Trivial solutions therefore remain a potential bias when using expectedness as an indicator for novelty.

5.5 Added value of applying the novelty metric in the field of interaction design and children

As indicated in the introduction Section a of this paper, research in the field of interaction design and children on the ideation effectiveness of participatory design methods is scarce. This work has made a small step in this direction. By applying a novelty metric on data from children, we have shown that it is theoretically and practically possible to apply measures for ideation effectiveness for design solutions generated by 8–10 year old children. The calculated novelty also shows that children create significantly more novel design solutions using a prototyping method after they have performed an NGT method.

Beside the indicated research on a realistic function for novelty, further research is required into how the capability of children to generate novel ideas relates to their developmental stage. Also, it is necessary to gain more insight in what skills they need to have acquired to be effectively involved in ideation methods. This research is not only limited to novelty, but extends to more effectiveness measures, like variety, quantity and quality (Shah et al., 2003).

6 Conclusion

This paper reviewed the application of the novelty metric as proposed in Shah et al. (2003) to compare design methods used to value the input of children in a design process. The metric was applied to data from a field study that compared the novelty of design solutions generated by children through two different methods, the Nominal Group Technique (NGT) and Prototyping.

Unfortunately we can safely conclude that the usefulness of the metric by Shah et al. (2003) is limited. A review of the novelty metric with extreme examples, revealed two fundamental problems: (1) the proposed metric only yields sensible results for comparing design solutions for an attribute, prohibiting a comparison between attributes, artefacts and methods; and (2) the metric decreases differences in novelty scores when working with large datasets, which is a confounding factor in analysing for statistically significant differences.

To overcome these problems we have proposed an alternative procedure that uses percentiles to only use the least frequent design solutions in the comparison. The alternative method yields interpretable results when comparing methods. Secondly, it is especially useful when using large datasets.

Comparing NGT and prototyping using the alternative metric lead to a significant difference in novelty for the design solutions. However, the study did show a significant interaction effect for the order: performing nominal brainstorming before prototyping leads to significant more novel design solutions than the reverse order. We hypothesize that the use of a brainstorming technique prior to a prototyping method, improves the accessibility of creative ideas in a prototyping session. In further development of the novelty metric we indicate that further research is required to better understand the relationship between frequency and novelty.

Finally we have applied the metric novelty in the field of interaction design with children. As such, we contributed insights from the field of ideation effectiveness to the field of interaction design and children. Further research should lead to insights into which ideation methods facilitate children best in generating novel design solutions, and how the quality and quantity compares to design solutions generated by adults.

References

- Al-khatib, B. A. (2012). The effect of using brainstorming strategy in developing creative problem solving skills among female students in Princess Alia University College. *American International Journal of Contemporary Research*, 2, 29–38.
- Bekker, M., Beusmans, J., Keyson, D., & Lloyd, P. (2003). Kid Reporter: a user requirements gathering technique for designing with children. *Interacting with Computers*, 15, 187–202.
- Brown, D. C. (2014). Problems with the calculation of Novelty metrics. In *Proceedings Design Creativity Workshop, 6th Int. Conf. on Design Computing and Cognition (DCC14)*, London, UK. p. Paper 5.
- Coskun, H., & Yilmaz, O. (2009). A new dynamical model of brainstorming: linear, nonlinear, continuous (simultaneous) and impulsive (sequential) cases. *Journal of Mathematical Psychology*, 53, 253–264.
- Druin, A. (1999). *The Role of Children in the Design of New Technology*. Report.
- Ferent, C., & Doboli, A. (2011). Measuring the uniqueness and variety of analog circuit design features. *Integration, the VLSI Journal*, 44, 39–50.
- Fern, E. F. (2001). *Advanced Focus Group Research*. Thousand Oaks, California: Sage Publications.
- Galindo-Rueda, F., & Millot, V. (2015). Measuring design and its Role in Innovation. *OECD Science, Technology and Industry Working Papers*, 2015, 1–51.
- Grace, K., & Maher, M. L. (2014). What to expect when You're expecting: the role of unexpectedness in computationally evaluating creativity. In *Proceedings of the Fifth International Conference on Computational Creativity* (pp. 120–128).
- Heary, C. M., & Hennessy, E. (2002). The use of focus group interviews in pediatric health care research. *Journal of Pediatric Psychology*, 27, 47–57.
- Hennessy, E., & Heary, C. (2004). Exploring children's views through Focus Groups. In S. M. Greene, & D. M. Hogan (Eds.), *Researching Children's Experiences: Approaches and Methods* (pp. 236–252), Saga. book chapter 13.
- Hernandez, N. V. A., Shah, J. J. B., & Smith, S. M. C. (2010). Understanding design ideation mechanisms through multilevel aligned empirical studies. *Design Studies*, 31, 382–410.
- Hou, W., Komlodi, A., Lutters, W. G., Boot, L., & Cotton, S. R. (2006). Girls don't waste time: pre-adolescent attitudes toward ICT. In *Proceedings of the Conference on Human Factors in Computing Systems 2006* (pp. 875–880).
- Langford, J. L., & McDonagh, D. (Eds.). (2003). *Focus Groups: Supporting Effective Product Development*. Taylor and Francis.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, 1, 476–490.
- López-Mesa, B. A. B., Mulet, E. A., Vidal, R. A., & Thompson, G. C. (2011). Effects of additional stimuli on idea-finding in design teams. *Journal of Engineering Design*, 22, 31–54.
- Markopoulos, P., Read, J. C., MacFarlane, S., Hoysniemi, J., & Hysniemi, J. (2008). *Evaluating Children's Interactive Products* (1 edition).. Burlington: Morgan Kaufmann Publishers.
- McGlynn, R. P., McGurk, D., Effland, V. S., Johll, N. L., & Harding, D. J. (2004). Brainstorming and task performance in groups constrained by evidence. *Organizational Behavior and Human Decision Processes*, 93, 75–87.
- Michalko, M. (2006). *Thinkertoys: a Handbook of Creative-thinking Techniques*. Ten Speed Press.

- Muller, M. J. (2003). Participatory Design: the third space in HCI. In A. Sears, & J. A. Jacko (Eds.), *The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Application* (2009 edition). (pp. 1051–1069). New York: Erlbaum, chapter 9.
- Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined metrics for measuring ideation effectiveness. *Design Studies*, 30, 737–743.
- Osborn, A. F. (1942). *How to Think Up*. McGraw-Hill.
- Pardo, S., Vetere, F., & Howard, S. (2005). Broadening stakeholder involvement in UCD: designers' perspectives on child-centred design. In *Proceedings of OZCHI 2005, CHISIG* (pp. 1–9). Canberra, Australia: ACM Press.
- Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies*, 32, 348–383.
- Shah, J. J., Vargaz-Hernandez, N., & Smith, S. M. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, 24, 111–134.
- Sluis-Thiescheffer, W., Bekker, T., & Eggen, B. (2007). Comparing early design methods for children. In *Proceedings of the 6th International Conference on Interaction Design and Children IDC 07, ACM, New York, NY, USA* (pp. 17–24).
- Sluis-Thiescheffer, R. J. W., Bekker, M. M., Eggen, J. H., Vermeeren, A. P. O. S., & De Ridder, H. (2011). Development and application of a framework for comparing early design methods for young children. *Interacting with Computers*, 23, 70–84.
- Stienstra, M. (2003). *Is Every Kid Having Fun?*. Ph.D. thesis. University of Twente. Enschede.
- Thang, B., Sluis-Thiescheffer, W., Bekker, T., Eggen, B., Vermeeren, A., & De Ridder, H. (2008). Comparing the creativity of children's design solutions based on expert assessment. In *Proceedings of the 7th International Conference on Interaction Design and Children IDC 08, ACM, New York, NY, USA* (pp. 266–273).
- Verhaegen, P. A., Peeters, J., Vandevenne, D., Dewulf, S., & Duflou, J. R. (2011). Effectiveness of the PAnDA ideation tool. *Procedia Engineering*, 9, 63–76.
- Vermeeren, A. (2009). *What's The Problem?: Studies on Identifying Usability Problems in User Tests*. Ph.D. thesis. Delft University of Technology.
- Wilson, J. O., Rosen, D., Nelson, B. A., & Yen, J. (2010). The effects of biological examples in idea generation. *Design Studies*, 31, 169–186.