INTERNATIONAL DESIGN CONFERENCE - DESIGN 2018
https://doi.org/10.21278/idc.2018.0119

# A-POSTERIORI NOVELTY ASSESSMENTS FOR SEQUENTIAL DESIGN SESSIONS

L. Fiorineschi, F. S. Frillici and F. Rotini

## Abstract
Among the various novelty metrics available in literature, those developed by Shah and Vargas-Hernandez are frequently used for academic purposes. More precisely, their “a-posteriori” version is often used for assessing ideation effectiveness of idea generation methods. We observed that when in presence of sequential idea generation sessions of the same task, the application of the recalled metric could lead to misleading results. In this paper, we argue about this problem, and we also provide useful suggestions for a correct “a-posteriori” assessment for sequential idea generation sessions.

*Keywords: design evaluation, evaluation, design creativity, novelty, novelty assessment*

## 1. Introduction
Today, creativity and innovation constitute the lighthouse for industry and for those academic disciplines involved in studying and supporting industrial processes, products and practitioners. Accordingly, several research activities exist about the identification of more comprehensive definitions of the interrelated concepts of creativity and innovation, as well as the concepts of newness and usefulness. Consequently, many definitions of creativity can be found in literature (Sarkar and Chakrabarti, 2008a), and for some of them the concept of novelty plays a crucial role (Sarkar and Chakrabarti, 2011; Corazza and Agnoli, 2016; Jagtap, 2016).
Moreover, aiming at supporting creativity and innovation, several contributions can be found in literature concerning studies about cognitive aspects, the development of new design methods and tools, and also about idea generation approaches and/or their mutual integration, e.g. (Shah et al., 2001; Howard et al., 2010; Frillici et al., 2015; Fiorineschi et al., 2018). Consequently, in order to evaluate and validate the effectiveness of their proposals, researchers need to "assess creativity" of single generated ideas or group of ideas. Accordingly, literature acknowledges several studies on the effectiveness of methodological proposals for idea generation, e.g. (Chulvi et al., 2012; Vargas-Hernandez et al., 2013).
The metric proposed by Shah et al. (2003a) is one of the most acknowledged and used by researchers interested in assessing groups of ideas, where "ideation effectiveness" is assessed by means of four distinct metrics, i.e. "quantity", "quality", "variety" and "novelty".
This paper is focused on the latter metric, in particular to the "a-posteriori" version, which is claimed to measure the unusualness or "uncommonness" of a given idea in relation to the specific group of other ideas generated in the same design session, by the same designer or group of designers. Such a peculiarity means that, according to Srivathsavai et al. (2010), the a-posteriori approach cannot be used for comparing an idea/concept with precedent design sessions or with marketed products.
However, while by definition of "a-posteriori" metrics cannot be used to assess novelty in relation to reference products, the possibility to consider ideas from precedent sessions deserves to be examined

more closely. Indeed, by interpreting the work of Shah et al. (2003a), if the same designer (or group of designers) performs sequential design sessions for the same design task or experiment, the metric foresees that each of the related outcomes should be assessed independently on the precedent ones. However, in this way the assessments are not reliable, because the outcomes produced by a designer (group of designers) in precedent sessions of the same task or experiment, certainly constitute a knowledge background for the designer her/himself (themselves). For example, if the attributes of an idea conceived in a generic "first" design session are widely used among the set of generated ideas, it is consequently scored with a mediocre novelty. However, if the same idea and its attributes are simply "reused" by the same designer only one time in a "second" design session of the same task, it will wrongly get the highest novelty score in that session. Consequently, it would be impossible to comprehensively evaluate how each single design session (and the related boundary conditions) actually affects the novelty of design outcomes. This could be a non-negligible problem that could hinder the adoption of the a-posteriori novelty assessment approaches, especially for certain experiments devoted to evaluate the effects of important phenomena such as the incubation (Sio and Ormerod, 2009; Tsenn et al., 2014) and the introduction of specific stimuli (Sarkar and Chakrabarti, 2008b; López-Mesa et al., 2011). Therefore, if a previously conceived idea is “reused”, and not actually conceived in the current design session of the same design task (or experiment), its novelty score should be assigned accordingly, also by taking into account the knowledge acquired during precedent activities performed for the same task or experiment. Nevertheless, it does not mean to substitute the metric with some "a-priori" ones (Shah et al., 2003b), because the latter refers to a (theoretically) generally valid and "static" universe of reference solutions, while in case of sequential sessions, the a-posteriori approach should still refer to a case-sensitive universe, but which evolves dynamically with sessions.

Therefore, the recalled applicability problem of the a-posteriori metric of Shah et al. is not trivial to solve. Accordingly, with this paper we aim at comprehensively introducing the recalled issue, to argue about it, and to propose useful suggestions for its resolution.

In order to reach the recalled objective, we structured the paper as it follows.

An introduction is reported in Section 2, about the examined metric, together with a short survey on its acknowledged variants. Section 3 comprehensively introduces the observed applicability problem, and reports some possible strategy for performing the assessment. In Section 4, the preferred strategy is highlighted and some motivations are reported about the indicated preference, while Section 5 reports discussions and possible future developments. Finally, Conclusions are reported in the last section.

## 2. Related work

Among the different ways to assess creativity, literature acknowledges some specific metrics to evaluate ideas, new products, or even people (Johnson et al., 2016). Especially concerning creativity evaluations of products and processes, it is possible to identify different metrics based on a variety of parameters like originality, usefulness, fluency, flexibility, etc. (Vargas-Hernandez et al., 2012). One of the most diffused is that of Shah et al. (2003a), which aims at measuring ideation effectiveness in terms of quantity, quality, variety and novelty of generated ideas. More precisely, they identified "quantity" as the number of generated ideas, "quality" as a measure of feasibility and compliance of ideas with the design requirements, "variety" as a measure of the explored solution space, and "novelty" as how unusual or unexpected is an idea in relation to other ones. Moreover, the same authors take into consideration the different stages characterizing a generic design process (i.e. conceptual design, embodiment design and detail design). Shah et al. (2003a) also proposed four well-known metrics, i.e. one for each of the above mentioned parameters. Variety and Quantity metrics are intended to be used for assessing group of ideas and cannot be applied for evaluating a single idea, while Quality and Novelty are intended to score each single idea (Shah et al., 2000), but the related mean values can be used to assess groups of ideas. For Variety assessment, the authors proposed to analyse the generated ideas in terms of four items, i.e. physical principles (PP), working principles (WP), embodiment (EMB) and details (DET) characterizing each implemented function. Subsequently, a hierarchical tree can be built, i.e. the so called "genealogy tree" (GT), where nodes represent the number of ideas adopting a specific item variant, and lines hierarchically connect nodes belonging to the different items. Moreover,

each level is characterized by a different weight, which is used to take into account the different impacts that the four items are supposed to have on the variety score.
For assessing quality, the authors assert that well acknowledged procedures like Pugh matrices (Pugh, 1991) can be used to evaluate the overall quality of a group of ideas. Therefore, both functional and non-functional requirements could be taken into account for assessing the quality of a given idea.
Concerning Novelty, Shah et al. (2003a) considered two different procedures, i.e. the "a-priori" and the "a-posteriori" assessment. For the first approach, it is necessary to define a universe of reference ideas for each function or attribute, in order to assign a specific novelty score to each examined ideas. Other a-priori approaches for assessing product novelty are present in literature, like that of Sarkar and Chakrabarti (2011), where a specific framework is used to assess the degree of novelty with respect to previous products. Also an improvement of the recalled method has been recently proposed for a more comprehensive novelty evaluation (Jagtap, 2016).
In the second approach proposed by Shah et al. (2003a), i.e. the "a-posteriori" one (SNM), the authors suggest to identify some recurring "key-attributes" among the set of ideas to be examined, and to find the different ways in which each attribute has been satisfied. Moreover, each identified key attribute is associated to a normalized weight, indicating its relative importance level. Once the key attributes have been identified and the related weight assigned, novelty of ideas is assessed by scoring the solutions generated for each key attribute, and then the scores are summed together by multiplying each of them by the related normalized weights. More precisely the novelty score for a given attribute $S_{ij}$ is calculated by Equation 1:

$$S_{ij} = 10 \times \frac{T_{ij}-C_{ij}}{T_{ij}} \tag{1}$$

Where $T_{ij}$ is the total number of solutions (or ideas) conceived for the key attribute $i$, and design stage $j$; and $C_{ij}$ is the number of the solution characterizing the idea to be assessed, for the same attribute $i$, and design stage $j$. Then, the overall novelty of each idea M is calculated by Equation 2:

$$M_{SNM} = \sum_{i=1}^{m} f_i \sum_{j=1}^{n} S_{ij} p_j \tag{2}$$

Where $f_i$ is the weight of the attribute $i$, $m$ is the number of attributes, $n$ is the number of design stages and $p_j$ is the weight assigned to the design stage $j$.
In the example of the "a-posteriori" assessment reported in Shah et al. (2003a), no particular problems or difficulties are mentioned, but some limitations to the recalled a-posteriori novelty assessment have been highlighted in literature, and also some variants have been proposed to partially overcome them.

## 2.1. Acknowledged limits of the a-posteriori metric of Shah et al.

Brown (2014) highlighted several issues to SNM, among which, the subjectivity of the identification of key attributes (as for the a-priori version), the need to identify weights for each attribute, the need to define a clear separation between conceptual and embodiment design descriptions.
Vargas-Hernandez et al. (2012) pointed that the metrics could be improved in terms of responsiveness and boundary cases. More precisely, responsiveness was intended as the capability of the metric to accurately reflect changes on the set of ideas and, accordingly, Peeters et al. (2010) observed that the application of SNM to large sets of ideas reduces the sensitivity of the metric. Instead, boundary cases refer to the application of the metric to particular cases with "extreme" scores.
As quoted in (Johnson et al., 2016), it has also been observed that the SNM is only sensitive to differences between concepts at the embodiment level, excluding more abstract differences.
Sluis-Thiescheffer et al. (2016) recently observed that the example in the paper of Shah et al. (2003a) refers to a particular case where for the set of 46 generated ideas, 46 solutions have been found for each attribute, implying that each of the considered ideas successfully satisfies each attribute. However, depending on the considered sample of designers, the design task, and the particular examined design stage, it is possible that some of the conceived solutions only partially satisfy the same number of attributes.
However, maybe one of the most acknowledged limits of the a-posteriori novelty metric of Shah et al. (2003) consists in representing a relative novelty or, in other words, a non-absolute novelty. Indeed, it

does not compare ideas from past idea generation sessions or to marketed products (Srivathsavai et al., 2010). However, as introduced in Section 1, for certain experiments it is still unclear if ideas from precedent sessions of the same design task should be considered. Nevertheless, as shown in the following paragraphs, in case of positive answer, it is currently unclear "how" to consider them.

## 2.2. Variants of SNM

Here in the following we report a short description about the variants of SNM that we found in our literature review, which have been proposed by scholars to overcome some of the limits introduced above.

Peeters et al. (2010) proposed to consider the three degrees of novelty introduced by Pahl et al. (2007), i.e. original, adaptive and variant. More precisely, they assert that an idea is "original" if is able to solve a problem at the PP level of the GT. Similarly, an idea is "adaptive" if solves the problem at the WP level and then "variant" if the problem is solved at the EMB level. Accordingly, they practically use PP, WP and EMB items in place of the SNM attributes, and uses the standardized set of weights proposed by Nelson et al. (2009) (10 for PP, 5 for WP and 2 for EMB), but continues to apply Equations 1 and 2.

The proposal of Vargas-Hernandez et al. (2012) still consider GT, but focuses the attention on the WP level, and proposes a new formula for calculating S. However, Johnson et al. (2016) observed that the recalled metric variant have some limitations, i.e. the need to describe ideas at the embodiment level, same scores for siblings at the embodiment level, and possible problems with extreme scoring conditions. According to the recalled and other limitations (among which, the lack of consideration for uncommonness of PPs and WPs), they proposed a new metric variant based on GT, where and additional level has been added upon the PP one, i.e. the "Strategy" level. Moreover, they proposed a new formula for calculating M.

Eventually, Sluis-Thiescheffer et al. (2016) proposed to assess novelty with a binary metric (novel or not novel), where an arbitrary threshold is considered to identify less frequent (novel) solutions. However, in light of the current improvements of SMN reviewed here, it is possible to assert that the problem of novelty assessments for sequential design sessions of a same experiment has not been examined, and then still deserves to be investigated appropriately.

# 3. Assessing novelty in sequential design sessions

Some different scenarios may present the need to assess novelty of ideas or design outcomes generated by sequential sessions. For example, in order to evaluate the effects derived from incubation, Shah et al. (2003b) carried out an experiment where two sequential design sessions were performed under specific conditions. In the recalled test, the four metrics of Shah et al. (2003a) were used to assess design outcomes, by applying them on the sets of ideas generated on both the two idea generation sessions. For the purposes of the recalled experiment, SNM has been successfully applied because of the particular arrangement of the test: a control group without incubation time between the two sessions, and an analysis group with a certain incubation time between the two sessions. More precisely, since the first session was the same for the two groups, any eventual "delta" observable on design outcomes was reasonably attributable to the effects of incubation. Therefore, it seems that SNM can be used in experiments with multiple design or idea generation sessions, but the assessment must be performed by considering a unique set of ideas for each of the examined groups. This could be quite limitative, since in this way it is not possible to comprehensively evaluate, for example, the effects derived from a second additional incubation time.

Another possible academic scenario involving multiple design or idea generation sessions, might be that where researchers are interested in evaluating the effects provided by the introduction of particular stimuli to the designers (Cardoso et al., 2009; Chiu and Shu, 2012; Cheng et al., 2014). Also in such case, the SNM metric can be used only for the evaluation of a single set of ideas for each designer (or group of designers), avoiding the evaluation of the "incremental" effects derived by the hypothetical sequential introduction of multiple stimuli. We faced this problem in our experimental activities, where we aimed at comparing the effects derived by the introduction of two different representations of a reference product, followed by the introduction of a particular verbal stimulus.

More precisely, we were interested in studying the effects provided by the representation sets belonging to two different conceptual design approaches (Pahl et al., 2007; Fiorineschi et al., 2016). Therefore, to evaluate the effects of the recalled representation schemes, we organized the testing procedure in three distinct idea generation sessions, where a sample of 35 students was subdivided in three groups (a control group and two analysis groups) and asked to perform the same specific design task. The first session was intended to provide initial and neutral information about the considered sample of convenience. Then, at the end of the first session, we collected the produced ideas and distributed the supplemental material to groups. More precisely, in this second session, each student received a paper sheet with the information structured according to the representation scheme assigned to the group (a "placebo" was assigned to the control group). Eventually, we also considered a third session to evaluate if different reactions were observable among the three groups, after the introduction of a verbal stimulus (the same for all groups) specifically inviting the participants to avoid fixation on a specific structural detail. The main characteristics of the recalled experiment are resumed in Table 1.

**Table 1. Experiment with sequential design sessions**

| | Group 1 | Group 2 | Group 3 |
|---|---|---|---|
| Session 1 | Same initial instructions to all groups | | |
| Session 2 | Graphical stimulus A | Graphical stimulus B | Graphical stimulus C |
| Session 3 | Same verbal stimulus to all groups | | |

For the novelty assessment, we decided not to use a-priori approaches, because we were unable to define a comprehensive set of reference products. Moreover, due to the high heterogeneity of the set of conceived ideas, for novelty assessments we opted for a particular SNM variant, basically inspired to the version proposed by Peeters et al. (2010). Nevertheless, the same considerations can be easily extended to SNM and all of its variants. Furthermore, the three sessions were identified as parts of the same design stage, then assigning j=1 to Equations 1 and 2.

However, in order to evaluate the incremental effects of the sequential sessions (and the related additional stimuli) on the novelty of generated ideas, two issues need to be faced: the calculation of the $S_{ij}$ scores (by Equation 1) for the calculation of the novelty of ideas (by Equation 2) and the calculation of the mean novelty.

The following paragraphs describe the possible strategies for facing the mentioned issues.

## 3.1. Strategies for the calculation of the S-scores

As shown by Equation 1 the value of the term $S_{ij}$ is strictly dependent on two terms, i.e. $T_{ij}$ and $C_{ij}$, whose values strictly depend on the selected set (or universe) of solutions. More precisely, for the experiment represented in Table 1, we evaluated three different but reasonably logic strategies:

- Strategy 1 (singular): For each group, in each session, consider only the set of ideas of the current session to be assessed.
- Strategy 2 (cumulative): For each group, consider the whole set of ideas from all of the three performed sessions.
- Strategy 3 (incremental): For each group, consider an expanding set of ideas, composed by:
    - Ideas from the first session when assessing the same first session
    - Ideas from the first and the second session when assessing the second session
    - Ideas from all of the three sessions when assessing the third session.

The three strategies listed above led to three different set of values for the $S_{ij}$ (that we calculated on the PP, WP and EMB items), which consequently led to three different set of novelty scores for assessed ideas. An excerpt of the performed assessment is reported in Tables 2, 3 and 4, respectively for the first, the second and the third strategy.

In the mentioned tables, it is possible to observe how the $S_{ij}$ scores for the same ideas can change, depending on the strategy adopted for performing the assessments. Accordingly, the same idea can get quite different novelty scores, mainly depending on the actual values that the terms $T_{ij}$ and $C_{ij}$ represent. For example, considering Idea 1 from Table 2, it is possible to observe that the related solution for the PP item is quite common in the first session, due to the low $S_{ij}$. However, it seems that it has been reused with a lower frequency in the successive sessions, leading to a higher $S_{ij}$ if calculated with the cumulative strategy (see Table 3). Differently, it seems that WP and EMB items of the same idea have been constantly reused, leading to a quite constant $S_{ij}$ value across the singular and the cumulative strategies.

**Table 2. Excerpt of the evaluation performed for the test introduced in Section 3, with the first strategy**

| Idea | Session | PP | WP | EMB | Novelty |
|---|---|---|---|---|---|
| 1 | 1 | 2,50 | 9,69 | 9,69 | 5,90 |
| 2 | 1 | 2,50 | 6,56 | 9,69 | 4,92 |
| 3 | 2 | 8,18 | 8,18 | 9,55 | 8,40 |
| 4 | 2 | 3,18 | 9,55 | 9,55 | 6,20 |
| 5 | 3 | 7,39 | 8,26 | 9,57 | 8,01 |
| 6 | 3 | 7,39 | 9,13 | 9,57 | 8,28 |

**Table 3. Excerpt of the evaluation performed for the test introduced in Section 3, with the second strategy**

| Idea | Session | PP | WP | EMB | Novelty |
|---|---|---|---|---|---|
| 1 | 1 | 4,16 | 9,87 | 9,87 | 6,86 |
| 2 | 1 | 4,16 | 7,92 | 9,87 | 6,25 |
| 3 | 2 | 8,31 | 8,57 | 9,87 | 8,64 |
| 4 | 2 | 4,16 | 9,87 | 9,87 | 6,86 |
| 5 | 3 | 8,31 | 8,57 | 9,87 | 8,64 |
| 6 | 3 | 4,16 | 7,92 | 8,96 | 6,10 |

**Table 4. Excerpt of the evaluation performed for the test introduced in Section 3, with the third strategy**

| Idea | Session | PP | WP | EMB | Novelty |
|---|---|---|---|---|---|
| 1 | 1 | 2,50 | 9,69 | 9,69 | 5,90 |
| 2 | 1 | 2,50 | 6,56 | 9,69 | 4,92 |
| 3 | 2 | 8,70 | 8,70 | 9,70 | 8,86 |
| 4 | 2 | 2,78 | 9,81 | 9,70 | 6,09 |
| 5 | 3 | 8,31 | 8,57 | 9,87 | 8,64 |
| 6 | 3 | 4,16 | 7,92 | 8,96 | 6,10 |

Considering Idea 6 (from the third session), the singular strategy led to a quite high novelty score (Table 1), while both the cumulative and incremental strategies revealed that especially the solution of the PP item is not so uncommon, since it has been used quite often in precedent sessions.

Different situations can be observed for Ideas 3 and 4 (from the second session), where it is possible to observe that in this cases, the three strategies led to three different sets of $S_{ij}$ values and then to three different sets of results for single-ideas novelty assessments.

As previously stated in this section, the same considerations can be extended to SNM and all the variants reviewed in Section 2. Indeed, although the evident differences, the fundamentals of the metrics are equally based on the comparison with the set of generated ideas.

### 3.2. Calculation of the mean novelty scores

As shown before, each of the strategies listed in Section 3.1 lead to different sets of novelty scores and then, it easy to infer that the mean values of the recalled sets will be different as well.

Obviously, there are different possibilities for the calculation of the recalled mean values, mainly dependent on the set of ideas (and related novelty scores) to be considered. Indeed, for each group (see Table 1) it could be possible to consider ideas from all of the three performed sessions, or from only the current session, or even from an incremental set, similarly to what happened for the incremental strategy in the $S_{ij}$ calculation. Then, it would be possible to combine each of the strategy for the $S_{ij}$ calculation with each of the mentioned strategies for the mean value calculation, leading to 9 possible different sets of mean novelty values for each group.

However, the considered kind of experiments are focused on the evaluation of the effects provided by each singular session. Therefore, among the 9 possibilities, the only ones that make sense (in terms of mean novelty calculation) are the three considering the novelty scores of ideas from the current session to be assessed. In other words, when assessing the mean novelty reached by a group in the first session of the experiment, only the set of novelty scores from the same session and the same group are considered. Similarly, for assessing the mean novelty reached by a group in the second session, only the set of novelty scores from the same second session are considered. The same for the third session.

As shown in Figure 1, Figure 2 and Figure 3 (neglecting statistical considerations, for the scope of this paper), the three strategies lead to different mean values and then to different observed trends across the three sessions.

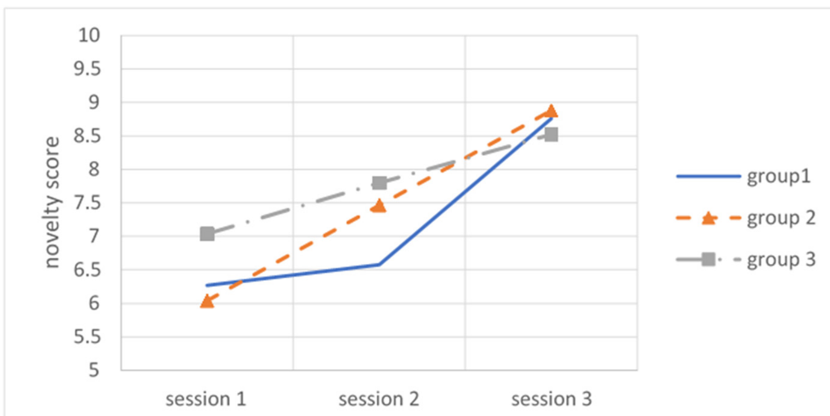


**Figure 1. Mean novelty scores calculated for each group for the set of novelty scores of ideas from the current session, calculated with the "singular" strategy**

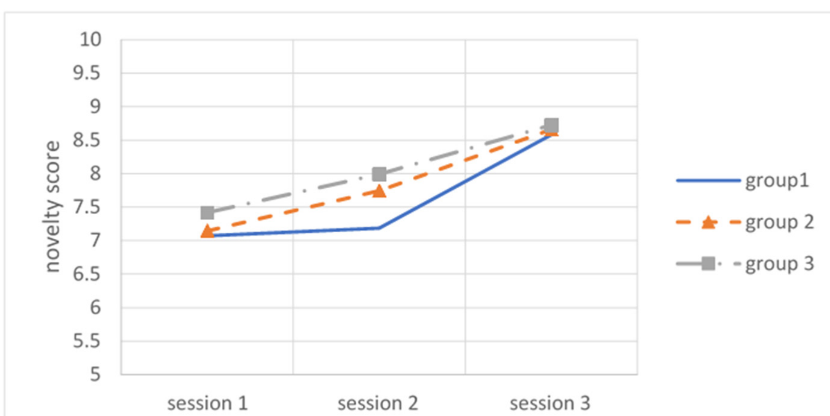


**Figure 2. Mean novelty scores calculated for each group for the set of novelty scores of ideas from the current session, calculated with the "cumulative" strategy**

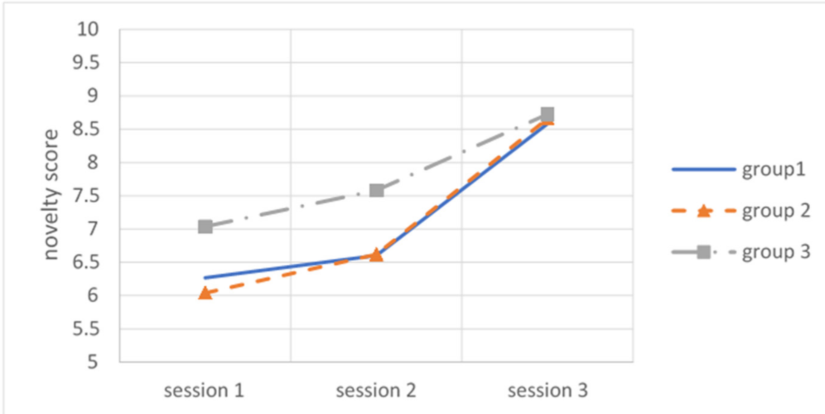


**Figure 3. Mean novelty scores calculated for each group for the set of novelty scores of ideas from the current session, calculated with the "incremental" strategy**

In particular, it is possible to observe that Figure 2 shows a flattened trend of the novelty scores reached by the three groups across the three sessions, while the other two strategies show a more sloped behaviour. Nevertheless, while the first strategy (Figure 1) shows different trends for the three groups, the third strategy (Figure 3) highlights that the first and the second groups behaved very similarly.

Therefore, in order to perform any kind of consideration about the experiment, it is necessary to select the best suited approach to achieve mean novelty scores.

## 4. Suggested assessment approach for sequential experiment sessions

As shown in the previous section, a-posteriori approaches for novelty assessment can be managed in a variety of ways if used for assessing design outcomes from sequential design or idea generation sessions. Therefore, in order to make any consideration about the outcomes of the performed experiment, it is crucial to select the right strategy for the calculation of novelty of single ideas and then for the mean novelty scores reached by the sample of designers. However, while it is easy to infer that for the type of the performed experiment the mean scores should be calculated only on the set of ideas from the current session, the selection of the strategy for the $S_{ij}$ calculation deserves to be performed with care.

Indeed, each of the different scores from the three approaches can provide useful information about the performed experiment. However, even if the first strategy (singular) could potentially allow to observe how novelty evolves across the three sessions for the three examined groups, the related novelty assessments for each idea after the first session are unreliable, as explained by the following extreme example. Suppose (ad absurdum) that a group produced exactly the same set of ideas in all the three sessions. In this case, by following the first strategy, the novelty scores of ideas are the same in each of the three phases. In our opinion, this is not correct since the ideas are simply "reused" across sessions, and not generated in the current ones. Reasoning in terms of $S_{ij}$ values, a solution for a given attribute or item that has been already used in precedent sessions by the same designer (or group of designers), cannot reach the same scores of those actually conceived in the current session and never used before.

In the second strategy (cumulative), the novelty of ideas of a generic session is calculated by referring to both precedent and future sessions (if present). For example, in the experiment represented in Table 1, ideas from the first session are assessed by referring to a universe composed by ideas generated also from the two subsequent sessions. However, it implies to compare novelty of ideas with "future" ideas, not yet conceived in the current session. It could make sense for evaluating the re-use of solutions actually conceived in the first and the second sessions, but their $S_{ij}$ values are not realistic, since calculated by referring to an expanded universe, where additional solutions (from the successive sessions) are considered.

The third strategy (incremental) is a sort of "limited" version of the cumulative approach, where ideas "ideas from the future" are neglected from the assessment. Accordingly, it can take into consideration the cumulative effect of ideas repetition across sequential sessions, but cannot be used to estimate the re-use of solutions in the successive sessions. More precisely, $T_{ij}$ should be considered as the total number of solutions conceived for the key attribute $i$, and design stage $j$, across all the currently performed design sessions. Similarly, $C_{ij}$ should be considered as the occurrences of the current solution for the key attribute $i$, in stage $j$, across all the currently performed design or idea generation sessions. Therefore, considering the experiment represented in Table 1, ideas from Session 1 can be normally assessed as indicated by SNM, i.e. $T_{ij}$ is the number of solutions for the key attribute (item) $i$ among the set of ideas conceived in the current session, and $C_{ij}$ represents the occurrences of the current solution for the key attribute $i$, in the same current session. Differently, when assessing ideas from the second session, $T_{ij}$ should be considered as the number of solutions for the key attribute (item) $i$, among the set of ideas conceived in the current session "plus" the set of ideas conceived in the precedent one. Similarly, $C_{ij}$ represents the occurrences of the current solution for the key attribute $i$, counted on both the first and the second sessions. Accordingly, in the third session, $T_{ij}$ should be considered as the number of solutions for the key attribute (item) $i$, among the set of ideas conceived in the current session "plus" the set of ideas conceived in both the first and the second sessions. Then, $C_{ij}$ represents here the occurrences of the current solution for the key attribute $i$, counted on the set of ideas generated across all the three sessions. Then, as shown in Tables 2, 3 and 4, this strategy led to the same results of the singular one for the first session, and the same results of the cumulative strategy in the third session. Differently, in the second session, the progressive strategy could take into consideration the effects of ideas reused from the first session, and reducing the $S_{ij}$ scores accordingly.

Therefore, among the possible strategies shown in Section 3.1, the incremental approach seems to be a good compromise, since it allows to show the effects of eventual re-use of ideas conceived in precedent sessions, and then to guarantee the assignment of highest scores only to solutions conceived in the current session. In this way, the effect of the boundary conditions characterizing the current session (e.g. incubation time or specific stimuli) can be evaluated with more precision.

To give an example, Table 5 reports the $T_{ij}$ and $C_{ij}$ counts of some solutions for the PP item, conceived by a group across the three sessions performed in the experiment represented in Table 1. Moreover, Table 6 reports the $S_{ij}$calculated by the three strategies introduced in Section 3.1.

**Table 5. $T_{ij}$ and $C_{ij}$ of three solutions related to the PP item, conceived by the first group of students across the three sessions**

| | First session | | Second session | | Third session | |
|---|---|---|---|---|---|---|
| | $C_{ij}$ | $T_{ij}$ | $C_{ij}$ | $T_{ij}$ | $C_{ij}$ | $T_{ij}$ |
| Fluid pressure | 24 | 32 | 15 | 22 | 6 | 23 |
| Capillarity | 3 | | 4 | | 6 | |
| Magnetism | 2 | | 0 | | 0 | |

Referring to Table 6, it is important to observe that while the singular and the incremental strategies provide a specific $S_{ij}$ value for each solution in each session, the cumulative strategy provides a single value, which is valid for all the three sessions. However, it is also important to observe that the "magnetism" solution appears only in the first session, and then the related $S_{ij}$ for the second and third sessions are not calculated by the first approach. Diversely, it is possible to observe how the $S_{ij}$ for "magnetism" increases session after session in the incremental strategy, starting from the same value calculated by the singular strategy in Session 1, up to the same value reached with the cumulative strategy in Session 3. This is quite obvious, because according to the fundamentals of the incremental strategy, in this case (no more "magnetism" solutions after Session 1) $C_{ij}$ remains unchanged after the first session, while $T_{ij}$ increases session after session. However, since ideas exploiting the magnetism solution for PP are present only in the first session, in this case the $S_{ij}$ values calculated for the second and third sessions by the incremental strategy are obviously not used in Equation 2.

**Table 6. $S_{ij}$ of three solutions of Table 5, calculated with the three different strategies (singular, cumulative and incremental)**

| | Singular | | | Cumulative | Incremental | | |
|---|---|---|---|---|---|---|---|
| | Sess. 1 | Sess. 2 | Sess. 3 | - | Sess. 1 | Sess. 2 | Sess. 3 |
| Fluid press. | 2,5 | 3,2 | 7,4 | 4,2 | 2,5 | 2,8 | 4,2 |
| Capillarity | 9,1 | 8,2 | 7,4 | 8,3 | 9,1 | 8,7 | 8,3 |
| Magnetism | 9,4 | - | - | 9,7 | 9,4 | 9,6 | 9,7 |

Nevertheless, what observed for the quite extreme example of the magnetism solution, reveals that the incremental strategy can actually keep track of the actual re-use of ideas across sessions. The effects of this peculiarity allow to reach more accurate evaluations. Considering the "capillarity" example of Tables 5, it is possible to observe that this solution is increasingly used session after session, but preserving relatively low counts of occurrences. Therefore, for the second and the third sessions, the incremental strategy lead to $S_{ij}$ values (Table 6) that are higher than those calculated by the "singular" strategy. Similarly, but in the opposed direction, the incremental strategy lead to lower $S_{ij}$ values for the "Fluid pressure" example, because although is rarely used in the third session, it has be used in more than an half of the ideas conceived in the first two sessions.

Therefore, for evaluating how "uncommonness" of ideas is influenced by the different triggers or conditions provided in sequential design or idea generation sessions, we suggest to adopt the incremental strategy for calculating $S_{ij}$ and novelty of single ideas. Moreover, in order to calculate the mean values for each group (or designer) and each session, we suggest to consider the scores of ideas conceived only the current session to be assessed. Accordingly, it is possible to assert that Figure 3 is the most representative of the experiment outcomes.

## 5. Discussions and future developments

The assessment approach suggested in this paper has been conceived to extend the usability of SNM and the related variants, to those cases where it is necessary to comprehensively map the contributions of multiple idea generation or design sessions.

Several implications can be ascribed to the results of our work, among which, the possibility to perform multi-session experiments, aimed at better understanding the effects derived by different and successive incubation intervals. Indeed, it is now possible to keep track of the novelty scores session by session, for both single ideas and group of ideas, and then highlighting the actual impact of the conceived incubation time. Moreover, as shown in the example represented in Table 1, similar considerations can be made for experiments devoted at investigating the effects of different sequences for providing specific stimuli. However, depending on the SNM version adopted for the assessment, the related problems still persists, and some of them should be carefully examined. Indeed, Peeters et al. (2010) observed that decrement of the SNM metric responsiveness when the set of ideas get more populated, and this is more or less valid for all of its variants. Unfortunately, since the proposed approach actually expands the reference universe of solutions, this certainly leads to a decreasing metric responsiveness, session after session. This could explain the "convergence" observed in Figure 3 for the third session.

Therefore, one of the possible future work could actually be focused on a comprehensive understanding the effects of the recalled limits. Moreover, the considerations made in this paper highlights that the plethora of novelty metrics available in literature, led to non-negligible difficulties in selecting the most suitable for the specific research and/or industrial needs. Therefore, this work can be considered a valid effort for a more comprehensive understanding of the potentialities of a-posteriori approaches, and the reported considerations could provide important information for the development of comprehensive "assessment guidelines". Lastly, the results presented here constitute a potential step for the development of a generally valid “a-posteriori” metric, allowing to be used for assessing the outcomes from (possibly any) generic design task.

## 6. Conclusions

The objective of this paper was to present a proposal capable to apply well-known a-posteriori novelty assessment approaches, also when in presence of sequential idea generation or design sessions. The performed review revealed that literature has not currently solved the issues characterizing the recalled cases, which we actually faced during our experimental activities. We used the results from the recalled experiment as example to explain the problem and to argue about it. Then, we proposed a particular strategy for considering "expanding" universes of solutions, session after session, in order to keep track of the evolving knowledge of the involved designers. Consequently, the new strategy allowed to comprehensively map the evolution of the novelty scores across multiple sessions, both for single ideas and for groups of ideas. In the latter case, we considered the mean values of the novelty scores reached by the solutions belonging to the current session to be examined.

Nevertheless, some potential limits of the proposals have been highlighted, especially about the decreasing responsiveness of the adopted metrics, when the set of solutions gets more populated session after session. Accordingly potential future works have been inferred, aimed at analysing the effects of the recalled limit and at contributing to the development of guidelines and comprehensive approaches for novelty assessment, applicable to any general experimental or industrial case.

## References


Brown, D.C. (2014), *Problems with the Calculation of Novelty Metrics*. [online] Available at: http://web.cs.wpi.edu/~dcb/Papers/DCC14/DCC14-Brown-Novelty-workshop.pdf

Cardoso, C., Badke-Schaub, P. and Luz, A. (2009), "Design Fixation on Non-Verbal Stimuli: the Influence of Simple vs Rich Pictorial Information on Design Problem-Solving", *Proceedings of ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, San Diego, California, USA, August 30 - September 2, 2009,* ASME, pp. 995-1002. https://doi.org/10.1115/DETC2009-86826

Cheng, P., Mugge, R. and Schoormans, J.P.L. (2014), "A new strategy to reduce design fixation: Presenting partial photographs to designers", *Design Studies*, Vol. 35 No. 4, pp. 374–391. https://doi.org/10.1016/j.destud.2014.02.004

Chiu, I. and Shu, L.H.H. (2012), "Investigating effects of oppositely related semantic stimuli on design concept creativity", *Journal of Engineering Design*, Vol. 23 No. 4, pp. 271–296. https://doi.org/10.1080/09544828.2011.603298

Chulvi, V., Mulet, E., Chakrabarti, A., López-Mesa, B. and González-Cruz, C. (2012), "Comparison of the degree of creativity in the design outcomes using different design methods", *Journal of Engineering Design*, Vol. 23 No. 4, pp. 241–269. https://doi.org/10.1080/09544828.2011.624501

Corazza, G.E. and Agnoli, S. (2016), *Creativity in the Twenty First Century Multidisciplinary Contributions to the Science of Creative Thinking*, Springer, Singapore. https://doi.org/10.1007/978-981-287-618-8

Fiorineschi, L., Frillici, F.S. and Rotini, F. (2018), "Enhancing functional decomposition and morphology with TRIZ: Literature review", *Computers in Industry*, Vol. 94, pp. 1–15. https://doi.org/10.1016/j.compind.2017.09.004

Fiorineschi, L., Rotini, F. and Rissone, P. (2016), "A new conceptual design approach for overcoming the flaws of functional decomposition and morphology", *Journal of Engineering Design*, Vol. 27 No. 7, pp. 438–468. https://doi.org/10.1080/09544828.2016.1160275

Frillici, F.S., Fiorineschi, L. and Cascini, G. (2015), "Linking TRIZ to conceptual design engineering approaches", *Procedia Engineering*, Vol. 131, pp. 1031-1040. https://doi.org/10.1016/j.proeng.2015.12.421

Howard, T.J., Dekoninck, E.A. and Culley, S.J. (2010), "The use of creative stimuli at early stages of industrial product innovation", *Research in Engineering Design*, Vol. 21 No. 4, pp. 263–274. https://doi.org/10.1007/s00163-010-0091-4

Jagtap, S. (2016), "Assessing Design Creativity: Refinements to the Novelty Assessment Method", *Proceedings of the DESIGN 2016 / 14th International Design Conference, Dubrovnik, Croatia, May 16-19, 2016*, The Design Society, Glasgow, pp. 1045–1054.

Johnson, T.A., Cheeley, A., Caldwell, B.W. and Green, M.G. (2016), "Comparison and Extension of Novelty Metrics for Problem-Solving Tasks", *Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Charlotte, USA, August 21-24, 2016*, ASME, pp. V007T06A012. https://doi.org/10.1115/DETC2016-60319

López-Mesa, B., Mulet, E., Vidal, R. and Thompson G. (2011), "Effects of additional stimuli on idea-finding in design teams", *Journal of Engineering Design*, Vol. 22 No. 1, pp. 31-54. https://doi.org/10.1080/09544820902911366

Nelson, B.A., Wilson, J.O., Rosen, D. and Yen, J. (2009), "Refined metrics for measuring ideation effectiveness", *Design Studies*, Vol. 30 No. 6, pp. 737–743. https://doi.org/10.1016/j.destud.2009.07.002

Pahl, G., Beitz, W., Feldhusen, J. and Grote, K.-H. (2007), *Engineering Design*, 3rd ed., Springer-Verlag, London. https://doi.org/10.1007/978-1-84628-319-2

Peeters, J., Verhaegen, P.-A., Vandervenne, D. and Duflou, J.R. (2010), "Refined Metrics for Measuring Novelty in Ideation", *Proceedings of IDMME - Virtual Concept 2010, Bordeaux, France, October 20-22, 2010*, Springer, pp. 1–4.

Pugh, S. (1991), *Total Design: Integrated Methods for Successful Product Engineering*, Addison-Wesley, Reading, Massachusetts.

Sarkar, P. and Chakrabarti, A. (2008a), "Studying engineering design creativity - Developing a Common Definition and Associated Measures", *Proceedings of the NSF International Workshop on Studying Design Creativity'08, Aix-en-Provence, France, March 10-11, 2008.*

Sarkar, P. and Chakrabarti, A. (2008b), "The effect of representation of triggers on design outcomes", *AI EDAM*, Vol. 22 No. 2, pp. 101–116. https://doi.org/10.1017/S0890060408000073

Sarkar, P. and Chakrabarti, A. (2011), "Assessing design creativity", *Design Studies*, Vol. 32 No. 4, pp. 348–383. https://doi.org/10.1016/j.destud.2011.01.002

Shah, J.J., Kulkarni, S.V. and Vargas-Hernandez, N. (2000), "Evaluation of Idea Generation Methods for Conceptual Design : Effectiveness Metrics and Design", *Journal of Mechanical Design*, Vol. 122 No. 4, pp. 377–384. https://doi.org/10.1115/1.1315592

Shah, J.J., Vargas-Hernandez, N. and Smith, S.M. (2003), "Empirical Studies of Design Ideation: Alignment of Design Experiments with Lab Experiments", *Proceedings of DETC 2003 / ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Chicago, Illinois, September 2-6, 2003*, ASME, pp. 847-856. https://doi.org/10.1115/DETC2003/DTM-48679

Shah, J.J., Vargas-Hernandez, N. and Smith, S.M. (2003), "Metrics for measuring ideation effectiveness", *Design Studies*, Vol. 24 No. 2, pp. 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

Shah, J.J., Vargas-Hernandez, N., Summers, J.D. and Kulkarni, S. (2001), "Collaborative Sketching (C-Sketch) - An Idea Generation Technique for Engineering Design", *The Journal of Creative Behavior*, Vol. 35 No. 3, pp. 168–198. https://doi.org/10.1002/j.2162-6057.2001.tb01045.x.

Sio, U.N. and Ormerod, T.C. (2009), "Does incubation enhance problem-solving?", *Psychological Bulletin*, Vol. 135 No. 1, pp. 94–120. https://doi.org/10.1037/a0014212

Sluis-Thiescheffer, W., Bekker, T., Eggen, B. and Vermeeren, A. (2016), "Measuring and comparing novelty for design solutions generated by young children through different design methods", *Design Studies*, Vol. 43, pp. 48–73. https://doi.org/10.1016/j.destud.2016.01.001

Srivathsavai, R., Genco, N., Hölttä-Otto, K. and Seepersad, C.C. (2010), "Study of Existing Metrics Used in Measurement of Ideation Effectiveness", *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2010, Montreal, Canada, August 15-18, 2010*, ASME, pp. 355-366. https://doi.org/10.1115/DETC2010-28802

Tsenn, J., Atilola, O., McAdams, D.A. and Linsey, J.S. (2014), "The effects of time and incubation on design concept generation", *Design Studies*, Vol. 35 No. 5, pp. 500–526. https://doi.org/10.1016/j.destud.2014.02.003

Vargas-Hernandez, N., Okudan, G.E. and Schmidt, L.C. (2012), "Effectiveness Metrics for Ideation: Merging Genealogy Trees and Improving Novelty Metric", *Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2012, Chicago, USA, August 12-15, 2012*, ASME, pp. 85-93. https://doi.org/10.1115/DETC2012-70295

Vargas-Hernandez, N., Schmidt, L.C. and Okudan, G.E. (2013), "Systematic Ideation Effectiveness Study of TRIZ", *Journal of Mechanical Design*, Vol. 135 No. 10, pp. 101009. https://doi.org/10.1115/1.4024976

Dr. Francesco Saverio Frillici, Research assistant
University of Florence, Industrial Engineering Department
via di Santa Marta, 3, 50139 Florence, Italy
Email: francescosaverio.frillici@unifi.it